

Der Einfluss von Testmotivation auf die Leistung in einem Leistungstest

Ines Giermann

Diplomarbeit als Teil der Diplomprüfung für Psychologen
und Psychologinnen

Universität Regensburg

24.01.2012

Inhaltsverzeichnis

Inhaltsverzeichnis.....	2
0. Einleitung.....	6
1. Leistungsmotivation, Intelligenz und Testmotivation	8
1.1 Leistungsmotiv und Leistungsmotivation.....	8
1.1.1 Begriffsbestimmung von Leistungsmotiv und Leistungsmotivation.....	9
1.1.2 Theorien der Leistungsmotivation	11
1.1.2.1 Das Risikowahlmodell nach Atkinson (1957)	11
1.1.2.2 Kritik am Risikowahlmodell und weiterführende Theorien	14
1.1.2.3 Das quadripolare Modell der Leistungsmotivation von Covington et al. (Covington & Omelich, 1991; Covington & Roberts, 1994)...	17
1.1.3 Leistungsmotivmessung.....	19
1.1.4 Leistungsmotivkomponenten.....	21
1.2 Intelligenz	24
1.2.1 Begriffsbestimmung von Intelligenz	24
1.2.2 Die Messung von Intelligenz	25
1.2.3 Strukturtheorien der Intelligenz	28
1.2.3.1 Die Zwei-Faktoren-Theorie von Spearman (1904)	29
1.2.3.2 Das Primärfaktorenmodell von Thurstone (1938).....	30
1.2.3.3 Fluide und kristalline Intelligenz nach Cattell (1963, 1971)	31
1.2.3.4 Das Drei-Ebenen-Modell nach Carroll (1993)	31
1.2.4 Intelligenz und Schulleistung.....	33
1.3 Testmotivation.....	35
1.3.1 Begriffsbestimmung von Testmotivation	36
1.3.2 Modell der Testmotivation.....	37
1.3.3 Empirische Befunde zur Testmotivation	41
1.3.3.1 Der Einfluss von Testmotivation auf die Testleistungen.....	41
1.3.3.2 Die Auswirkungen von Testmotivation in Leistungstests	43
1.4 Testmotivation, Leistungsmotiv und Leistungstests.....	45
1.4.1 Zusammenführung der verschiedenen Leistungsmotive in das quadripolare Modell der Leistungsmotivation.....	45
1.4.2 Leistungsmotiv und Leistungstests.....	46
1.4.3 Der Zusammenhang zwischen Testmotivation, Leistungsmotiv und Leistungstests.....	47

2. Hypothesen und Fragestellungen	50
2.1 Leistungsmotiv	50
2.2 Testmotivation.....	51
2.3 Intelligenz	53
2.4 Zusammenhänge	53
2.4.1 Leistungsmotiv und Testmotivation	54
2.4.2 Testmotivation und Leistungen in einem Leistungstest	55
2.4.3 Schulleistung, besuchte Schulart bzw. Jahrgangsstufe und Intelligenz.....	56
3. Methode	58
3.1 Beschreibung der Stichprobe.....	58
3.1.1 Besuchte Schulart und Klassen.....	59
3.1.2 Geschlecht nach Schule und Klasse.....	60
3.1.3 Alter	60
3.1.4 Schulfächer und Schulnoten	61
3.1.5 Klasse wiederholt.....	62
3.2 Untersuchungsablauf.....	63
3.3 Statistische Datenauswertung – verwendete Verfahren.....	65
3.4 Untersuchungsmethode	67
3.4.1 Strukturgleichungsmodell	68
3.4.2 Einverständniserklärung, soziodemographischer Fragebogen und Ergebnisrückmeldung	70
3.4.3 Das Regensburger Leistungs-Motiv-Inventar für Kinder und Jugendliche (RLMI-K/J)	70
3.4.3.1 Beschreibung	71
3.4.3.2 Testanalyse.....	71
3.4.4 Das Leistungsmotivgitter für Kinder und Jugendliche – Kurzversion (AMG-S K-J)	73
3.4.4.1 Beschreibung	74
3.4.4.2 Testanalyse.....	75
3.4.5 Fragebogen zur Erfassung der Testmotivation	76
3.4.5.1 Beschreibung	77
3.4.5.2 Testanalyse.....	79
3.4.5.3 Konfirmatorische Faktorenanalyse	81

3.4.6	Das neue Leistungsprüfsystem (LPS-neu).....	86
3.4.6.1	Beschreibung	86
3.4.6.2	Testanalyse.....	90
3.5	Prüfung auf Unterschiedlichkeit zwischen verschiedenen Gruppen	94
4.	Ergebnisse.....	101
4.1	Leistungsmotiv	101
4.2	Testmotivation.....	107
4.3	Intelligenz	111
4.4	Zusammenhänge	115
4.4.1	Leistungsmotiv und Testmotivation	115
4.4.2	Testmotivation und Leistungen in einem Leistungstest	119
4.4.3	Schulleistung, besuchte Schularart bzw. Jahrgangsstufe und Intelligenz.....	128
5.	Diskussion	144
5.1	Diskussion der Methode	144
5.1.1	Untersuchungsmethode.....	144
5.1.1.1	Das Regensburger Leistungs-Motiv-Inventar für Kinder und Jugendliche (RLMI-K/J).....	144
5.1.1.2	Das Leistungsmotivgitter für Kinder und Jugendliche – Kurzversion (AMG-S K-J)	145
5.1.1.3	Das neue Leistungsprüfsystem (LPS-neu).....	145
5.1.1.4	Fragebogen zur Erfassung der Testmotivation	147
5.1.1.5	Fragebogen zu den soziodemographischen Daten.....	148
5.1.2	Stichprobe	148
5.1.3	Untersuchungsablauf	150
5.2	Diskussion der Ergebnisse.....	151
5.2.1	Leistungsmotiv.....	152
5.2.2	Testmotivation	153
5.2.3	Intelligenz	154
5.2.4	Zusammenhänge	155
5.2.4.1	Leistungsmotiv und Testmotivation	155
5.2.4.2	Testmotivation und Leistungen in einem Leistungstest	157
5.2.4.3	Schulleistung, besuchte Schularart bzw. Jahrgangsstufe und Intelligenz	158
5.3	Nachträgliche Hypothesen	161

0. Einleitung

Schulleistungsstudien sorgen in Deutschland immer wieder für große Aufregung, nicht nur die bekannteste unter ihnen, die PISA-Studie. Kritik am Schulsystem wurde laut, aber auch an der Untersuchungsmethode selbst: Die Testergebnisse spiegeln eher die Motivation als die Leistungsfähigkeit der Schüler/-innen wider, und das werde von den zuständigen Forschungsinstituten häufig nicht berücksichtigt.

In der Psychologie werden die meisten Erkenntnisse durch Testungen beziehungsweise Untersuchungen mit Patienten oder anderen Versuchspersonen ermöglicht. Sie alle nehmen freiwillig teil, doch wie sieht es mit der Motivation der Testpersonen aus? Sind sie ausreichend motiviert oder sogar übermotiviert und ist das nicht ein Punkt, der bei solchen Testungen beachtet werden sollte? „Bedeutet ein großes Interesse an einem möglichst hohen Ergebnis gleichzeitig auch, dass das Testergebnis der maximalen Leistungsfähigkeit entspricht?“ (Bossong, 1985, S. 88) Und haben nicht Versuchspersonen, bei denen es meistens um Testungen ohne Konsequenzen geht, eine andere Motivation als beispielsweise Patienten, die durch die Erkenntnisse solcher Tests vielleicht sogar gesund werden können?

Ziel dieser Untersuchung soll es sein, die Testmotivation, also die Motivation, in einer bestimmten Situation oder in einem Test erfolgreich zu sein (Eklöf, 2008),

1. genauer zu untersuchen
2. sowie der Frage nachzugehen, wie die Testmotivation erhoben werden kann
3. und welche Auswirkungen diese auf die Leistung in einem Leistungstest hat, der ohne Konsequenzen für die Versuchspersonen bleibt.

Dabei dient der neu überarbeitete Intelligenztest, das Leistungsprüfsystem (LPS-neu) von Kreuzpointner (2010), als Test ohne Konsequenzen.

Detailliertes Wissen über diese Zusammenhänge könnte nicht nur in der Intelligenzforschung und in dem schulischen Bereich von Nutzen sein, vor allem bei Tests die ohne Konsequenzen für die Schüler/-innen bleiben, sondern auch für andere psychologische Gebiete Bedeutung haben. Denn nicht nur bei der Interpretation von Intelligenztestergebnissen sollte die Testmotivation eine Rolle spielen. Sie sollte auch bei anderen psychologischen Testungen und deren Interpretation Beachtung finden sowie darüber hinaus bereits bei der Gestaltung von Tests und bei den Testinstruktionen berücksichtigt werden, um sicher zu sein, dass ein Test nicht nur den Mangel an Testmotivation erfasst, sondern tatsächlich das zu messende Konstrukt bestimmt.

Im ersten Kapitel werden die Theorien und bisherigen Erkenntnisse zu den Bereichen Leistungsmotivation – da die Testmotivation eine situative Komponente dieser ist –, Intelligenz und Testmotivation sowie die Zusammenhänge zwischen den Konstrukten vorgestellt. Dieses führt schließlich zu den Fragestellungen und Hypothesen des zweiten Kapitels. Im dritten Kapitel werden das methodische Vorgehen dieser Studie, die Stichprobe sowie die einzelnen Erhebungsinstrumente vorgestellt. Im anschließenden Kapitel (Kapitel 4) werden die empirischen Ergebnisse mithilfe von Grafiken und Tabellen beschrieben, wobei diese im fünften Kapitel diskutiert werden. Im letzten Kapitel werden die wichtigsten Ergebnisse dieser Untersuchung nochmals zusammengefasst dargestellt.

1. Leistungsmotivation, Intelligenz und Testmotivation

1.1 Leistungsmotiv und Leistungsmotivation

Um das Konstrukt des Leistungsmotivs und der Leistungsmotivation besser verstehen zu können, wird zunächst geklärt, was mit den Begriffen Motiv und Motivation gemeint ist und womit man sich in dieser Teildisziplin der Psychologie beschäftigt. Festgehalten werden sollte jedoch die Tatsache, dass es viele verschiedene Begriffsbestimmungen zu Motiven und Motivationen gibt. Diese reichen vom Alltagsverständnis, darüber, was motiviert heißt, bis hin zu den Anfängen, den triebtheoretischen Konzepten, wie beispielsweise das von Hull, welcher das Streben nach einem Ziel als Motivation betrachtete (Atkinson, 1975), oder den instinkttheoretischen und ethologischen Konzepten, zum Beispiel von Darwin (Schneider & Schmalt, 2000). Aufgrund dieser vielen Aspekte kann die nachfolgende Begriffsbestimmung von Motiv und Motivation auch nur einen Versuch darstellen, diese einzugrenzen.

Die Ausdrücke Motiv und Motivation lassen sich auf das lateinische Wort „movere“ zurückführen, was im Deutschen bewegen beziehungsweise antreiben heißt. Dabei sind Motive zeitlich relativ stabile Vorlieben einer Person, welche eine Neigung darstellen, gewisse Themen unterschiedlich, positiv oder negativ, zu bewerten (Keller, 1981; Rheinberg, 2006). Im Rahmen der Motivationspsychologie wird der Standpunkt vertreten, dass Motive für das Handeln einer Person verantwortlich sind, in Abhängigkeit von ihren Bewertungsvorlieben (Rheinberg, 2006; Vollmeyer & Brunstein, 2005), was wiederum bedeutet, dass sie von Person zu Person verschieden sind. Demnach wird durch Motive zu erklären versucht, weshalb sich Menschen verhalten, wie sie sich verhalten (Schmalt & Meyer, 1976) und was sie dazu veranlasst, ein Verhalten in einer bestimmten Häufigkeit zu zeigen oder nicht zu zeigen (Keller, 1981). Bei den Motiven können vor allem die drei Motive – Anschluss-, Macht- und Leistungsmotiv – unterschieden werden (Vollmeyer & Brunstein, 2005). In dieser Arbeit wird jedoch nur auf das Leistungsmotiv eingegangen.

Rheinberg (2006) definiert Motivation als „die aktivierende Ausrichtung des momentanen Lebensvollzugs auf einen positiv bewerteten Zielzustand“ (S. 15). Diese Definition beinhaltet zugleich auch eine Abwendung von einem Zielzustand, der als zumindest relativ negativ bewertet wird. Dementsprechend ist nach Rheinberg (2006) ein Verhalten dann motiviert, wenn die Handlung ein Ziel hat. Jedoch wird die dabei gegenwärtig vorhandene Motivation nicht nur von personenbezogenen Einflüssen wie den Motiven,

Bedürfnissen oder Zielen geprägt, sondern ist auch von den situationsbezogenen Faktoren, den Anreizen, abhängig (Vollmeyer & Brunstein, 2005). Ob eine Person also motiviert ist, ein Ziel zu verfolgen und ein entsprechendes Verhalten zu zeigen, ist sowohl von den Anreizen der Situation, den Präferenzen der Person als auch von deren Wechselwirkungen abhängig (Rheinberg, 2006). Nach Schmalt (1986) sollte der Motivationsbegriff erklären können, welche Vorgänge im Organismus ablaufen, während eine Handlung begonnen wird, während sie aufrechterhalten und dann beendet wird.

Im Zuge der Motivationsforschung wird demnach erklärt, wie ein Verhalten aufgenommen und beendet wird, welche Ausrichtung das Verhalten abhängig vom Ziel hatte und wie das Verhalten gesteuert, aufrechterhalten und energetisiert wird (Schmalt, 1986). Dabei wird selbstverständlich angenommen, dass Menschen durch praktische Erfahrungen, ihre Veranlagungen, auf eine gewisse Art und Weise zu handeln, verändern können. Aus diesem Grund beschäftigen sich die Motivationsforscher auch damit, „wie etwas in der Vergangenheit Gelerntes und etwas momentan vom Individuum Wahrgenommenes zusammen mit noch anderen Faktoren die Richtung, Stärke und Persistenz seines Verhaltens bei einer bestimmten Gelegenheit beeinflussen“ (Atkinson, 1964/1975, S. 18). Zusammenfassend lässt sich festhalten, dass durch die Motivationsforschung versucht wird, die Richtung und die Ausdauer, aber auch die Intensität von Verhalten zu erklären (Rheinberg, 2006).

1.1.1 Begriffsbestimmung von Leistungsmotiv und Leistungsmotivation

Im folgenden Kapitel soll nun explizit auf das Leistungsmotiv und die Leistungsmotivation eingegangen werden. Aufgrund der langen Forschungstradition ist das Leistungsmotiv das bis heute am besten untersuchte Motiv (Brunstein & Heckhausen, 2006) beziehungsweise die Leistungsmotivation ist „die mit Abstand besterforschte Klasse von Person-Umweltbezügen“ (Rheinberg, 2006, S. 59). Bereits 1938 wurde der Begriff des Leistungsmotivs durch Murray geprägt, welcher einen Motivkatalog erstellte, in dem er Bedürfnisse und Motive in Kategorien einteilte. In diesem definierte er das Leistungsbedürfnis, „need for achievement“, wie folgt:

To accomplish something difficult. To master, manipulate or organize physical objects, human beings or ideas. To do this as rapidly and independently as possible. To overcome obstacles and attain a high standard. To excel one's self. To rival and surpass others. To increase self-regard by the successful exercise of talent. (zitiert nach Heckhausen, 1989, S. 231)

Obwohl Murray mit dieser Definition einen wichtigen Beitrag leistete, gelten McClelland et al. (1953) und Atkinson (1957, 1958) als die Begründer der Leistungsmotivationsforschung. So definierten McClelland und Kollegen (1953) ein Verhalten, welches leistungsmotiviert ist, als die Auseinandersetzung „with some standard of excellence“ (S. 110). Folglich liegt ein leistungsmotiviertes Verhalten nur dann vor, wenn es darauf abzielt, die eigene Tüchtigkeit zu bewerten und es sich mit der Erreichung oder Übertreffung eines selbstgesetzten Gütemaßstabes auseinandersetzt. Der Handelnde möchte in einem speziellen Aufgabenfeld, unabhängig davon, ob es sich um wichtige und folgenreiche oder um objektiv unwichtige Dinge handelt, erfahren, was er gerade noch schaffen kann und was nicht mehr (Rheinberg, 2006, 2004). Dabei spielen Aufgaben, bei denen die Möglichkeit besteht, gleichermaßen Erfolg und Misserfolg zu erleben, eine entscheidende Rolle. Aufgrund dessen fügte Atkinson (1957) diese beiden Möglichkeiten in sein Verhaltensmodell ein und trennte zudem die Situationsvariablen von den überdauernden Variablen. Als Situationsvariablen beschrieb er die Erwartungen von Erfolg und Misserfolg sowie die Anreize derselben. Die beiden Personenvariablen, Erfolg aufzusuchen und Misserfolg zu vermeiden, sind die Motive, die Atkinson als Erfolgs- und Misserfolgsmotiv bezeichnete (Schmalt & Meyer, 1976).

Beim Leistungsmotiv handelt es sich um eine überdauernde Verhaltensdisposition, Erfolg aufzusuchen und Misserfolg zu vermeiden (Schmalt & Meyer, 1976), welche von Mensch zu Mensch und von Situation zu Situation verschieden ist (Rheinberg, 2006). Aus diesem Grund versuchen Leistungsmotivationsforscher Erklärungen zu liefern, weshalb bestimmte Aufgaben und Aktivitäten ausgewählt werden, wie ausdauernd sich Menschen bei deren Durchführung zeigen und welche Leistungen sie letztendlich darin erbringen (Eccles, Wigfield & Schiefele, 1998).

Leistungsmotivation liegt dementsprechend dann vor, wenn eine Person Leistungsziele willentlich anstrebt und dazu das Leistungsmotiv durch Hinweisreize aktiviert wird, die von der jeweiligen Situation abhängig sind. Dabei ist sowohl das Verhalten als auch das Erleben der Person auf die Zielzustände, also das Erleben von Erfolg oder Misserfolg, ausgerichtet. Zudem muss beim Vorliegen der Leistungsmotivation der Antrieb für eine Handlung von der Person selbst ausgehen und sie sollte sich gleichfalls einem Tüchtigkeitsmaßstab verpflichtet fühlen (Brunstein & Heckhausen, 2006).

1.1.2 Theorien der Leistungsmotivation

Es gibt viele verschiedene Theorien, die das Thema des leistungsmotivierten Verhaltens behandeln. Die Anfänge dieser Theorien gehen auf Murray sowie auf die Arbeiten von McClelland et al. (Schmalt & Meyer, 1976) und Atkinson zurück. Bei McClelland et al. stand die affektbezogene Leistungsmotivationsforschung im Vordergrund. Einen anderen Ausgangspunkt wählt Atkinson (1957) für das Risikowahlmodell, in welchem er vornehmlich die kognitiven, vor allem situationsspezifischen Prozesse hervorhebt. Dieses Modell soll wegen seiner auch heute noch großen Wichtigkeit im Feld der Leistungsmotivationsforschung im Folgenden vorgestellt werden. Aufgrund der am Risikowahlmodell geübten Kritik wurden, ausgehend von diesem Modell, viele weitere Theorien entwickelt. Hier soll nur eine kleine Auswahl der daraus entstandenen Theorien, die Attributionstheorie von Weiner et al. (1971), das Selbstbegründigungsmodell der Leistungsmotivation von Heckhausen (1972, 1975) sowie das quadripolare Modell der Leistungsmotivation von Covington et al. (Covington & Omelich, 1991; Covington & Roberts, 1994) genauer vorgestellt werden.

1.1.2.1 Das „Risikowahlmodell“ von Atkinson (1957)

Das Risikowahlmodell von Atkinson basiert auf den Modellüberlegungen der sogenannten Erwartungs-mal-Wert-Theorien „von TOLMAN (1932) und LEWIN und Mitarbeitern (LEWIN et al., 1944), in denen Verhalten von den beiden Grundparametern ‚Erwartung‘ und ‚Wert‘ beziehungsweise ‚Nutzen‘ eines Ziels abhängig gemacht wird“ (Schmalt & Meyer, 1976, S. 13). Im Rahmen dieses Modells wird versucht, leistungsmotiviertes Verhalten als eine Funktion von Situations- und Personenvariablen zu erklären (siehe Kapitel 1.1.1). Nach Lewin (1938) handelt es sich bei Leistungssituationen demnach um einen klassischen Appetenz-Aversions-Konflikt. Von diesem Konflikt wird gesprochen, wenn eine Person Situationen gegenübersteht, die sowohl positive als auch negative Konsequenzen haben. Ein veranschaulichendes Beispiel für diesen Konflikt wäre ein Kind, das einen Hund zwar streicheln möchte, gleichzeitig jedoch Angst davor hat, gebissen zu werden (Rheinberg, 2006). Personen stellen sich Leistungssituationen, um Erfolge zu erzielen, wobei jedoch auch Misserfolge, die vermieden werden sollen, möglich sind.

Gemäß der Kernannahme des Risikowahlmodells entspricht Leistungsmotivation beziehungsweise die Zuwendung oder Meidung einer Leistungsaufgabe der Resultierenden Tendenz (RT). Diese wiederum ergibt sich dabei aus der Summe von aufsuchender Tendenz (Te), also der Hoffnung auf Erfolg, und meidender Tendenz (Tm), der Furcht vor Misserfolg (die negativ ist):

$$RT = Te + Tm$$

(Schmalt, 1986; Schneider, Gallitz & Meise, 1973; Weiner, 1994). Die Resultierende Tendenz ist somit die Kraft, die ein Verhalten initiiert und dieses aufrechterhält. Des Weiteren geht Atkinson in seinem Modell davon aus, dass in einer Leistungssituation immer beide motivationalen Tendenzen angesprochen werden, und zwar gleichzeitig. Welche Tendenz überwiegt, ist davon abhängig, welches dieser überdauernden Motive stärker ausgeprägt ist. Zu jeder der motivationalen Teilsysteme zählt eine Kraft, die den Handelnden in gegensätzliche Richtung leiten will (Schmalt, 1986). Die beiden Motivkomponenten Te, die motivationale Tendenz, Erfolg aufzusuchen, und Tm, die motivationale Tendenz, Misserfolg zu meiden, setzen sich dabei wie folgt zusammen:

$$Te = Me \cdot Ae \cdot We$$

$$Tm = Mm \cdot Am \cdot Wm$$

(Atkinson, 1964/1975). Die Tendenz, Erfolg anzustreben (Te) bzw. Misserfolg zu vermeiden (Tm), wird als Produkt von Motiv, Erfolgsmotiv (Me) bzw. Misserfolgsmotiv (Mm), Anreiz des Erfolgs (Ae) bzw. des Misserfolgs (Am) und der subjektiven Wahrscheinlichkeit, Erfolg (We) bzw. Misserfolg (Wm) zu haben, angesehen (Atkinson, 1964/1975).

Unter einem Motiv versteht Atkinson (1957) ganz allgemein das Streben nach einer bestimmten Art von Befriedigung. Dabei ist das misserfolgsmeidende Motiv dadurch gekennzeichnet, auf einen Misserfolg mit Scham und Verlegenheit zu reagieren (Atkinson, 1964/1975). Im Gegensatz dazu kann das erfolgsaufsuchende Motiv im Falle eines Erfolges, mit dem Gefühl des Stolzes zu reagieren, beschrieben werden (Atkinson, 1957). Erwartung ist für Atkinson (1957) eine kognitive Vorwegnahme des Ausgangs einer Handlung, welche durch situative Hinweisreize angeregt wird. Sie wird durch die individuelle und subjektive Erfolgswahrscheinlichkeit (We) beziehungsweise Misserfolgswahrscheinlichkeit (Wm) repräsentiert. Der Anreiz eines Zieles, also die relative Attraktivität bzw. Unattraktivität desselben, wird im Risikowahlmodell „als lineare Funktion der subjektiven Erfolgswahrscheinlichkeit: $Ae = 1 - We$ und $Am = We$ “ (Schneider et al., 1973, S. 19) definiert. Demzufolge steigt die Attraktivität einer Auf-

gabe mit der Schwierigkeit. Ist die Aufgabe schwierig, ist der Erfolg groß, wenn man sie gemeistert hat. Ist die Aufgabe leicht, so trifft ein Misserfolg umso härter.

Infolge der linearen, inversen Annahme der Beziehung von Anreiz und Erwartung werden die motivationalen Tendenzen, Erfolg aufzusuchen bzw. Misserfolg zu vermeiden, bei einer mittleren Erfolgswahrscheinlichkeit von $We = .50$ maximal (Schneider et al., 1973). Folglich sollten Personen, bei denen die erfolgsaufsuchende Tendenz überwiegt, Aufgaben mit einer Erfolgswahrscheinlichkeit von 50 % bevorzugen und zudem ihre Ausdauer und Anstrengung in diesem Bereich maximieren. Misserfolgsmeidende Personen sollten in diesem Schwierigkeitsbereich sowohl in ihrer Leistung als auch in ihrer Ausdauer maximal gehemmt sein.

Die Überprüfung der Vorhersagen des Modells bezüglich der Aufgabenwahl wurde mithilfe des Ringwurfexperiments (Schmalt, 1986; Weiner, 1994) in zahlreichen Untersuchungen, unter anderem von Atkinson und Litwin (1960), von Hamilton (1974) und von Schneider und Meise (1973) (zitiert nach Schmalt, 1986), durchgeführt. In diesen Experimenten sollten die Versuchspersonen über einen Pflock mehrmals nacheinander einen Ring werfen, wobei sie die Entfernung zum Pflock selbst bestimmen konnten. Dabei nahmen Atkinson und Litwin (1960) an, dass die Erfolgswahrscheinlichkeit nahe am Pflock positiv ist und mit steigender Distanz dazu abnimmt. Die Entfernung, welche die Probanden aussuchen, stellt somit einen Verhaltensindikator für die Aufgabenwahl und die Aufgabenschwierigkeit dar (Weiner, 1994). In dem Experiment von Hamilton konnte gezeigt werden, dass die deutlichsten Unterschiede der Wahlhäufigkeiten zwischen Erfolgs- und Misserfolgsmotivierten im mittleren Schwierigkeitsbereich liegen (Schmalt, 1986). Laut der Theorie sollten die Misserfolgsmotivierten auf beide Extreme der Skala der Erfolgswahrscheinlichkeit ausweichen, bei Hamilton hingegen bevorzugten sie nur die sehr schwierigen Aufgaben (Schmalt, 1986). Auch andere Untersuchungen konnten zeigen, dass Misserfolgsmotivierte ihre Auswahl über die gesamte Schwierigkeitsskala streuen, nur auf ein Ende der Skala ausweichen oder sich wie die Erfolgsmotivierten verhielten, wenn auch nicht in der gleichen ausgeprägten Form (Schmalt, 1986).

Bei Schneider et al. (1973) handelten die Misserfolgsmotivierten allerdings theoriekonform, jedoch nur dann, wenn sich die Auswahl auf objektive Schwierigkeitsstufen bezog. Bei subjektiven Erfolgswahrscheinlichkeiten der Auswahl geht dieses Befundmuster verloren (Schmalt, 1986). Dass Misserfolgsmotivierte immer nur sehr schwere oder sehr leichte Aufgaben wählen, konnte nur bei Aufgaben gefunden werden, die für

sie persönlich relevant waren und somit einen Einfluss auf das Selbstbewusstsein haben können. Ist das nicht der Fall, dann wählen sie andere Aufgaben (Schneider et al., 1973).

Es wurden zahlreiche weitere Studien zur Überprüfung des Risikowahlmodells durchgeführt, die diese Theorie zumindest teilweise belegen konnten (vgl. hierzu Heckhausen & Heckhausen, 2006).

1.1.2.2 Kritik am Risikowahlmodell und weiterführende Theorien

Das Risikowahlmodell wurde aufgrund der empirischen Untersuchungen oft kritisiert und teilweise revidiert, wobei hier nur ein kleiner Auszug der Kritik am Modell sowie Erweiterungen der Theorie vorgestellt werden kann.

Nach Schneider und Mitarbeitern (1973) wurde Kritik vor allem an der Annahme geübt, dass Anreiz und Erfolgserwartung linear, invers zusammenhängen. So weisen beispielsweise Schneider et al. (1973) und Heckhausen (1963) darauf hin, dass Erfolgsmotivierte zwar meist die mittleren Erfolgserwartungen bevorzugen, sie jedoch diejenige Schwierigkeitsstufe präferieren, bei der jeder dritte Versuch zum Erfolg führt und nicht, wie nach dem Modell zu vermuten wäre, jeder zweite Versuch. Zudem wurde Kritik dahingehend geäußert, dass die Erfolgswahrscheinlichkeit ausschließlich vom Anreiz abhängt. So sind für den Anreizwert einer Aufgabe unter anderem auch die zeitliche Distanz zum Ziel und die persönliche Bedeutung des Ziels wichtig, wobei der Anreiz zudem durch die Verfolgung weiterer Ziele mitbestimmt wird (Schneider et al., 1973). Aus diesem Grund sollte nach Schneider und Mitarbeitern (1973) die Definition des Anreizes um den Punkt „persönlich relevant“ erweitert werden. Zudem sollten bei Aufgaben, die weniger bedeutsam sind, Strategien aufgedeckt werden, die die Misserfolgsmotivierten verwenden. Darüber hinaus wird kritisiert, dass Aufgabeneigenschaften wie die Komplexität einer Aufgabe im Modell nicht berücksichtigt werden. Diese sollte nach Schneider et al. (1973) unter anderem zusammen mit den Anreizqualitäten für die Motivation in das Modell mit eingeschlossen werden. Nur so können Aussagen für den Erfolg und Misserfolg auf die Leistung getroffen werden.

Aufgrund der zahlreichen empirischen Untersuchungen und deren Ergebnissen wurde das Risikowahlmodell häufig in Teilbereichen revidiert (Schmalt & Meyer, 1976). So wurde zunächst unter Beibehaltung der Grundparameter versucht, das Modell zu erweitern. In den Erweiterungen wurden verhaltensdeterminierende Größen der psycho-

logischen Vergangenheit wie zurückliegende Erfolge und Misserfolge und die psychologische Zukunft hinzugenommen (Schmalt & Meyer, 1976). Eine weitere Modellerweiterung fand durch die Einführung der Trägheitstendenz statt. Diese Trägheitstendenz wird aber nicht nur bei der gleichen Handlung, die misslungen oder unterbrochen wurde und wieder aufgenommen wird, wirksam, sondern bei allen Aktivitäten, die folgen und zu dem gleichen Motiv gehören (Schneider et al., 1973). Folglich führt eine motivationale Tendenz, die bei Erfolgsmotivierten – es überwiegt die Aufsuchen-Tendenz – nicht befriedigt wird, in einer darauffolgenden Leistungssituation zu einer stärkeren Anstrengung. Bei Misserfolgsmotivierten hingegen sollte eine Tendenz zum Verlassen der Leistungssituation bzw. eine Verschlechterung der Leistung vorliegen, da bei ihnen die misserfolgsmeidende Motivation nach einem Misserfolg größer wird. Die nachfolgenden Studien zur Trägheitstendenz, vor allem von Weiner, wiesen jedoch in der experimentellen Realisation Schwächen auf (Schneider et al., 1973).

Wie es dazu kommt, dass sich eine Person in bestimmten Situationen so und nicht anders verhält, blieb durch das Risikowahlmodell ebenfalls unberücksichtigt. Auf Grundlage der von Heider (1958) vorgeschlagenen Ursachenzuschreibungen für das eigene Abschneiden bei Aufgaben entwickelten Weiner und Kollegen (1971) ihre Attributionstheorie. Dabei schlagen sie ein Klassifikationsschema vor, in welchem die verwendeten Kausalfaktoren sowohl nach Stabilitätsfaktoren (stabil, variabel) als auch danach eingestuft werden, ob es sich um internale, also die Person betreffende, Faktoren oder um externale, die Umwelt betreffende, Faktoren handelt (Rheinberg, 2006). Die Theorie geht nun davon aus, dass die Erwartung Einfluss auf die Stabilitätsdimension nimmt und die Lokationsdimension Einfluss auf den Anreiz. Aus diesem Grund müssten die Kausalattributionen, also die Ursachenzuschreibungen für Erfolg und Misserfolg, die Motivationskonsequenzen eines Leistungsergebnisses beeinflussen (Weiner et al., 1971). Es zeigte sich, dass Erfolgs- und Misserfolgsmotivierte, wenn es die Situationsbedingungen zuließen, sich in ihren Kausalattributionen unterscheiden. So überwiegt bei den Erfolgsmotivierten die Tendenz, Erfolg internal zu attribuieren, diesen also vor allem den eigenen Fähigkeiten zuzuschreiben. Misserfolg hingegen schreiben sie insbesondere zeitvariablen Faktoren, wie beispielsweise mangelnder Anstrengung, zu. Diese Zuschreibung bewirkt, dass sie sich Leistungssituationen gerne aussetzen, aufgrund hoher Anstrengung Erfolg haben, was sich wiederum positiv auf den Selbstwert auswirkt (Rheinberg, 2006). Misserfolgsmotivierte hingegen schreiben Erfolge meist dem Glück oder der Aufgabenleichtigkeit zu und attribuieren Misserfolg auf den Mangel der

eigenen Fähigkeiten. So kann es in Leistungssituationen zu einer verringerten Anstrengung oder gar zu einer Vermeidung solcher Situationen kommen. Durch die Attributionstheorie konnte geklärt werden, warum sich Personen aufgrund ihrer Ursachenzuschreibung in bestimmten Situationen so und nicht anders verhalten. Jedoch wurde bei dieser Theorie übersehen, dass die Ursachenerklärung bei der Selbstbewertung nur ein vermittelnder Prozess ist (Rheinberg, 2006).

Nicht nur kognitive, sondern auch affektive Komponenten werden in dem Selbstbewertungsmodell von Heckhausen (1972, 1975) berücksichtigt. Er stützt sein Modell sowohl auf das Risikowahlmodell als auch auf die Attributionstheorie. Dabei besteht es aus drei Teilprozessen, die wechselseitig aufeinander einwirken. Es handelt sich bei diesen Teilprozessen um den Vergleich eines Ergebnisses mit einem Gütemaßstab, um die Kausalattribution des Ergebnisses und um den Affekt der Selbstbewertung, wie zufrieden beziehungsweise unzufrieden man mit der eigenen Tüchtigkeit ist. So bewerten, laut diesem Modell, erfolgsmotivierte Personen die Anforderungen realistisch und wählen bevorzugt Aufgaben, bei denen sie ihre Fähigkeiten mithilfe eines Gütemaßstabes messen können. Bei Erfolgen sind erfolgsmotivierte Personen stolz auf ihre eigenen Fähigkeiten und stabilisieren dabei ihren positiven Selbstbewertungsaffekt. Misserfolge haben aufgrund der zeitvariablen externalen Attribuierung auf zukünftige Leistungssituationen kaum Auswirkungen (Rheinberg, 2006). Im Gegensatz dazu meiden Misserfolgsmotivierte realistische Anforderungen, sodass Erfolge mit zeitvariablen externalen Faktoren erklärt werden und in künftigen Leistungssituationen auf die Erfolgserwartung kaum Einfluss nehmen. Misserfolge haben hingegen einen negativen Einfluss auf die Erfolgserwartung in zukünftigen Leistungssituationen, da diese auf die zeitstabilen internalen Faktoren attribuiert werden. Als Selbstbewertungsaffekt zeigt sich oft Scham, was sich wiederum negativ auf die Motivation auswirkt (Rheinberg, 2006). Im Gegensatz zu anderen Modellen der Leistungsmotivation wurde das Selbstbewertungsmodell bis heute, trotz der theoretischen Weiterentwicklungen auf dem Gebiet der Leistungsmotivationsforschung, nicht revidiert, weshalb die Annahmen weiterhin Gültigkeit besitzen (Rheinberg, 2006).

1.1.2.3 Das quadripolare Modell der Leistungsmotivation von Covington et al. (Covington & Omelich, 1991; Covington & Roberts, 1994)

Covington und Kollegen (Covington & Omelich, 1991; Covington & Roberts, 1994) haben eine Theorie der Leistungsmotivation entwickelt, die sich zunächst an dem Risikowahlmodell von Atkinson (1964/1975) orientiert, indem sie ebenfalls annehmen, dass die Leistungsmotivation durch die zwei Motive, Erfolgs- und Misserfolgsorientierung, gekennzeichnet ist. Die Theorie unterscheidet sich jedoch auch von Atkinsons Modell, da Covington et al. (Covington & Omelich, 1991; Covington & Roberts, 1994) davon ausgehen, dass sich die Komponenten der Leistungsmotivation als zwei unabhängige Dimensionen darstellen lassen und nicht, wie Atkinson (1964/1975) vorgeschlagen hatte, als eine einzige bipolare Dimension, die er durch die Bildung eines Differenzwertes der beiden Motive erhielt. Zudem nehmen Covington et al. (Covington & Omelich, 1991; Covington & Roberts, 1994) eine Unterscheidung der Leistungsmotivation in vier verschiedene Typen von Leistungsmotivierten vor, weshalb auch vom quadripolaren Modell der Leistungsmotivation gesprochen wird. Eine Darstellung der nach diesem Modell zu unterscheidenden Leistungsmotivtypen kann der Abbildung 1.1 entnommen werden. Da das genannte Modell für diese Arbeit als Basis der Leistungsmotivation dient, wird es hier noch genauer vorgestellt.

Covington und Kollegen (Covington & Omelich, 1991; Covington & Roberts, 1994) betrachten in ihrem Modell Erfolgs- und Misserfolgsorientierung als zwei unabhängige Merkmale der Leistungsmotivation und stützen sich dabei auf den Befund, dass diese Merkmale bei Selbstberichten nur mittelmäßig hohe Zusammenhänge aufweisen (Brunstein & Heckhausen, 2006). Außerdem können diese beiden Motive bei einer Person in jeder möglichen Ausprägung auftreten.

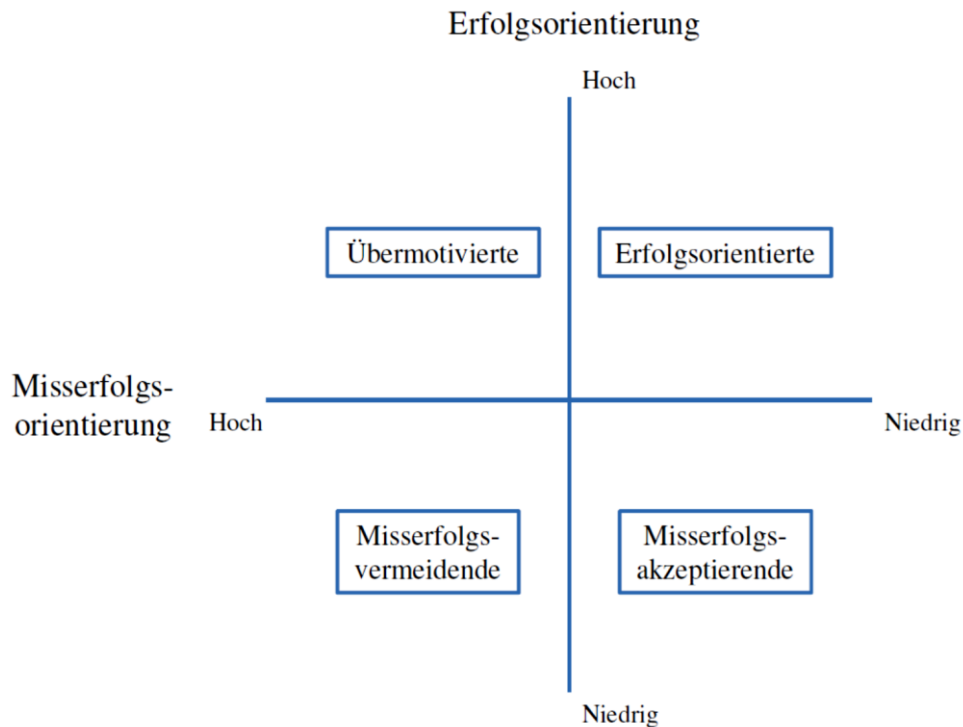


Abbildung 1.1

Quadrupolares Modell der Leistungsmotivation (Covington & Robert, 1994, S. 160; zitiert nach Brunstein & Heckhausen, 2006, S. 177)

Wie der Abbildung 1.1 entnommen werden kann, unterscheiden Covington et al. (Covington & Omelich, 1991; Covington & Roberts, 1994) je nach Ausprägung der Motive die vier Leistungsmotivtypen: Erfolgsorientierte, Misserfolgsvermeidende, Übermotivierte und Misserfolgsakzeptierende. Dabei entsprechen die beiden Typen Erfolgsorientierte, gekennzeichnet durch das Streben nach Erfolg und wenig Angst vor Misserfolg (HL = high hope for success and low fear for failure), und Misserfolgs-vermeidende, gekennzeichnet durch wenig Erfolgszuversicht und hohe Misserfolgs-angst (LH = low hope for success and high fear for failure) (Covington & Omelich, 1991), den Motivkomponenten von Atkinson (1964/1975). Zusätzlich nehmen Covington und Omelich (1991) noch weitere Unterscheidungen vor, nämlich in den Typ der Übermotivierten, geprägt durch das Streben nach Erfolg und einer hohen Angst vor Misserfolg (HH = high hope for success and high fear for failure), und in den Typ der Misserfolgsakzeptierenden, welche weder nach Erfolg streben noch vor Misserfolgen Angst haben (LL = low hope for success and low fear for failure). Diese beiden Komponenten werden nun noch etwas genauer beleuchtet, da sie sich nach Covington und Roberts (1994) in verschiedenen Aspekten voneinander unterscheiden.

Übermotivierte weisen sowohl in der Erfolgs- als auch in der Misserfolgsorientierung hohe Werte auf. Sie arbeiten hart und es ist ihnen wichtig, erfolgreich zu sein; dabei haben sie aber gleichzeitig auch große Angst vor einem Misserfolg (Covington & Omelich, 1991). Zudem sind ihnen leistungsbezogene Situationen wichtig, doch setzen sie sich oft selbst unter Druck und können bei Erfolg nur selten Stolz oder Zufriedenheit verspüren (Covington & Roberts, 1994). Sind Erfolgsorientierte meist neugierig und stellen sich selbstbewusst neuen Aufgaben, so sind Übermotivierte oft planlos oder verlieren sich im Detail, weshalb sie in Testsituationen häufig ineffektiv bleiben (Covington & Roberts, 1994).

Im Gegensatz zu den Übermotivierten weisen die Misserfolgsakzeptierenden sowohl in der Erfolgs- als auch in der Misserfolgsorientierung geringe Werte auf. Aufgrund der Abkoppelung ihres Selbstwertgefühls von gesellschaftlich angesehenen Leistungen empfinden sie weder Scham bei Misserfolg noch weisen sie besondere Anstrengung auf, um ihre Fertigkeiten zu verbessern (Covington & Roberts, 1994). Im Unterschied zu Misserfolgsmeidenden zeigen sie sich in Leistungssituationen wenig besorgt.

Zwar verweisen Brunstein und Heckhausen (2006) eher auf den illustrierenden als auf den erklärenden Wert der Befunde des quadripolaren Modells, dennoch besteht für sie ein Vorteil dieses Modells in der Abgrenzung von Personen, für die Motivkonflikte ein zentraler Bestandteil des Leistungsverhaltens sind, also den Übermotivierten, von den Personen, für die Leistungssituationen keinen Anreiz besitzen, den Misserfolgsakzeptierenden. Zudem erhöht sich durch die Behandlung der Leistungsmotivation als zwei unabhängige Dimensionen, der Informationswert, da sowohl der Leistungsmotivtyp des Übermotivierten als auch der des Misserfolgsakzeptierenden nach Subtraktion des Misserfolgsmotivs vom Erfolgsmotiv, wie in der traditionellen Leistungsmotivationsforschung gängig, den gleichen neutralen Wert erhalten würden (Brunstein & Heckhausen, 2006) und dadurch keine Aussagen über diese Motive möglich wären.

1.1.3 Leistungsmotivmessung

In der Literatur zur Leistungsmotivation finden sich eine Vielzahl von Verfahren, die das Leistungsmotiv zu messen versuchen. Hier sollen nur einige dieser Verfahren, angefangen mit den projektiven Messverfahren, vorgestellt werden.

Bereits 1938 entwickelte Murray den Thematischen Apperzeptionstest (TAT) als projektives Testverfahren. Mit diesem Verfahren wollte er implizite Motive als über-

dauernde Personenmerkmale messen (Rudolph, 2009). Nach Murray (1938) können diese zum Beispiel dadurch sichtbar gemacht werden, ob eine Situation gemieden bzw. aufgesucht wird oder wie eine Person mit Erfolg bzw. Misserfolg umgeht. Um diese Neigungen zu zeigen, müssen die Personen mittels Projektionen in die jeweiligen Situationen gebracht werden. So werden den Probanden beim TAT kurz Bildvorlagen gezeigt, zu denen sie anschließend eine Geschichte verfassen sollen (Rheinberg, 2004). Mithilfe eines Auswertungsschlüssels kann nach der Testdurchführung festgestellt werden, wie sehr eine Geschichte, ein bestimmtes Motiv angesprochen hat.

Mit dieser ersten Version konnte das Leistungsmotiv jedoch nicht ausschließlich und verlässlich gemessen werden, sodass McClelland et al. (1953) ein TAT-Verfahren zur Messung des Leistungsmotivs entwickelten, welches dieses eigenständig und standardisiert misst. In diesem Verfahren werden Bilder verwendet, die aufgrund leistungsthematischer Inhalte das Leistungsmotiv anregen und sich in den dazu verfassten Geschichten wiederfinden sollen. Mithilfe einer Untersuchung konnten McClelland und Kollegen (1953) einen standardisierten Inhaltsschlüssel entwickeln, durch den die Auswertung der verfassten Geschichten erfolgt sowie die Leistungsmotivation berechnet werden kann. In dieser Untersuchung mussten die Probanden die TAT-Geschichten entweder in einer entspannten, in einer neutralen oder in einer, unmittelbar vor dem TAT angeregten, leistungsthematischen Bedingung bearbeiten. Es zeigte sich, dass vor allem die Geschichten der Personen in der leistungsorientierten Bedingung und durch unmittelbar zuvor erlebter Erfolge oder Misserfolge stärker von leistungsthematischen Inhalten geprägt waren. Durch dieses Ergebnis war es McClelland und Kollegen (1953) letztendlich möglich, den standardisierten Inhaltsschlüssel zu entwickeln.

Eine zusätzliche Erweiterung des TAT-Verfahrens findet sich bei Heckhausen (1963). Dieser entwickelte eine Methode, in der nicht wie bei McClelland et al. (1953) das Erfolgs- und Misserfolgsmotiv zu einem Punktwert zusammengefasst wird, sondern die beiden Motivkomponenten zunächst gemeinsam erhoben werden und anschließend – mithilfe eines Auswertungsschlüssels – getrennt voneinander ausgewertet werden.

Allerdings ist der TAT jahrzehntelang kontrovers diskutiert worden. So besteht nach Rudolph (2009) ein gravierender Nachteil in der aufwendigen Auswertung, die nur von geschulten Personen durchgeführt werden kann. Zudem wird auch an der mäßigen Zuverlässigkeit und der Anfälligkeit für aktuelle Situationseinflüsse Kritik geübt. Wird der TAT jedoch im Zusammenwirken mit beispielsweise Verhaltensdaten eingesetzt, kann die Verwendung durchaus gerechtfertigt sein (Rheinberg, 2004).

Schmalt (1976, 1999) entwickelte mit seinem Leistungsmotiv-Gitter (LM-Gitter) eine einfachere und ökonomischere Methode als die des TAT-Verfahrens. Die in seinem Verfahren verwendete Gitter-Technik ist dem Verfahren des TAT jedoch ähnlich. So erhalten die Probanden Bilder und darunter Aussagen zu den Bildern. Die Aufgabe der Probanden besteht darin, diejenigen Aussagen zu wählen, die für sie am besten zu den Bildern passen. Es wird dabei angenommen, dass die Auswahl durch die Motive einer Person beeinflusst wird (Langens, Schmalt & Sokolowski, 2005). Mit dem vorgegebenen Antwortschema (siehe Kapitel 3.4.4) soll sowohl die Objektivität als auch die implizite Erhebung des Leistungsmotivs gewährleistet sein. Das in dieser Studie verwendete LM-Gitter für Kinder und Jugendliche in der Kurzversion (AMG-S K-J, Schmalt, 1999, 2005) teilt, wie auch die anderen Versionen des LM-Gitters, das Leistungsmotiv in drei Motivkomponenten ein (siehe Kapitel 1.1.4): Hoffnung auf Erfolg (HE), Furcht vor Misserfolg aktiv (FMa) und Furcht vor Misserfolg passiv (FMp).

Ein weiteres, von Lukesch und Peters-Häderle (2007) entwickeltes Messverfahren zur Erfassung des Leistungsmotivs ist das Regensburger Leistungsmotiv-Inventar für Erwachsene (RLMI-E). Auch dieses Verfahren beinhaltet ein vorgegebenes Antwortschema (siehe Kapitel 3.4.3), ähnlich dem des LM-Gitters von Schmalt. Im Gegensatz zum LM-Gitter werden den Probanden im RLMI-E zur Anregung des Leistungsmotivs leistungsthematische Situationen beschrieben. Mithilfe des RLMI-E lassen sich die beiden Leistungsmotivkomponenten HE und Furcht vor Misserfolg (FM) sowie zwei zusätzliche Komponenten – Furcht vor Erfolg (FE) und Hoffnung auf Misserfolg (HM) – messen (siehe Kapitel 1.1.4). Auch für dieses Verfahren wurde eine Version für Kinder und Jugendliche, das Regensburger Leistungsmotiv-Inventar für Kinder und Jugendliche (RLMI-K/J; Keil, 2009; Sowa, 2009; Schmidbauer, 2008; Lübke, 2007; Peters-Häderle, 2006) entworfen, welches ebenfalls in dieser Studie verwendet wird.

1.1.4 *Leistungsmotivkomponenten*

Es gibt verschiedene Auffassungen darüber, aus welchen Komponenten sich das Leistungsmotiv zusammensetzt. So wird nach der klassischen Motivationspsychologie, die auf die Erwartungs-mal-Wert-Theorie von Atkinson (1958) zurückgeht, lediglich zwischen den Komponenten Hoffnung auf Erfolg und Furcht vor Misserfolg unterschieden. Hinzu kommen aber weitere Überlegungen, die im Folgenden ebenfalls kurz dargestellt werden sollen.

Personen mit einer starken Hoffnung auf Erfolg wollen die eigene Tüchtigkeit sowie ihr eigenes Leistungsniveau steigern, indem sie sich mit Gütestandards auseinandersetzen und diverse Aufgaben gut oder noch besser machen (Brunstein & Heckhausen, 2006). Dabei können bei der Definition eines Gütemaßstabes nach Rheinberg (2006) zwei unterschiedliche Bezugsnormen, die individuelle (Personen vergleichen die Leistung mit ihrem bisherigen eigenen Leistungsniveau) und die soziale (Personen orientieren sich an Leistungen von Bezugsgruppen) angewandt werden. Personen mit einer starken Hoffnung auf Erfolg orientieren sich bei leistungsthematischen Aufgaben an der individuellen Bezugsnorm, weshalb ihnen der Vergleich mit anderen eher unwichtig ist (Langens et al., 2005). Hinzu kommt, dass erfolgsoptimistische Personen durch die Antizipation von Stolz bei erfolgreicher Bewältigung der Aufgabe motiviert werden und dieses beim Erreichen bzw. beim Übertreffen eines Gütemaßstabes auch empfinden (Brunstein & Heckhausen, 2006). Es überwiegt dementsprechend der positive Anreiz eines Erfolges, was zur Folge hat, dass sie sich leistungsthematischen Aufgaben gerne stellen (Rheinberg, 2004). Wichtig ist zudem, dass sich der Handelnde für das Ergebnis verantwortlich fühlt, Erfolg und Misserfolg gleich wahrscheinlich sind und er weiß, wann eine Aufgabe erfolgreich war (Atkinson, 1964/1975).

Personen, bei denen die Furcht vor Misserfolg überwiegt, vermeiden, wenn möglich, die Auseinandersetzung mit Gütemaßstäben, da die Bedrohung durch einen eventuellen Misserfolg größer als die Freude beim erfolgreichen Bestehen ist, auch wenn beide Ausgänge gleich wahrscheinlich sind (Atkinson, 1957; Heckhausen, 1963). Hinzu kommt, dass misserfolgsmeidende Personen durch die Antizipation von Scham beim Scheitern einer Aufgabe motiviert werden, diese zu meiden, da das ein eindeutiger Hinweis auf mangelnde Fähigkeiten wäre (Langens et al., 2005). Können sie Leistungssituationen aus wichtigen Gründen, wie zum Beispiel den Schulabschluss, jedoch nicht vermeiden, empfinden sie diese Aktivitäten als stark belastend (Rheinberg, 2006). Dabei kann es unter Umständen vorkommen, dass sie die Bearbeitung dieser Aktivitäten so weit hinausschieben, sodass sie aufgrund von Zeitmangel nicht mehr in der Lage sind, eine gute Leistung zu erbringen und sich somit für den Misserfolg nicht verantwortlich fühlen müssen (Langens et al., 2005). In Leistungssituationen, in denen sie sich die Aufgabenschwierigkeit aussuchen können, wählen sie meist Aufgaben, die keine Rückschlüsse auf den Leistungsstand sowie die eigene Tüchtigkeit erlauben, um kein Schamgefühl bei Misserfolg erleben zu müssen (Rheinberg, 2004).

Zusätzlich zu diesen beiden klassischen Leistungsmotiven teilten Schmalt und Kollegen (1976, 2005) aufgrund von Analysen der Ergebnisse des LM-Gitters das Misserfolgsmotiv in FMp und FMa ein. Beide Formen sind dadurch gekennzeichnet, dass sie einen Misserfolg in Leistungssituationen vermeiden wollen. Unterschiede finden sich hingegen auf der Erlebens- und Verhaltensorberfläche. So versuchen Personen mit einer hohen FMa den Misserfolg durch eine erhöhte Anstrengung zu vermeiden, wohingegen Personen mit einer hohen FMp meist inaktiv bleiben und versuchen, sich Leistungssituationen zu entziehen, um so den drohenden Misserfolg zu vermeiden (Schmalt, 2005). Bestätigung dieser Konzeption fand Schmalt (1976, 1999, 2005) in Untersuchungen an Schüler/-innen. Hier war die FMa mit positiven Leistungsmaßen verknüpft, sodass Schüler/-innen mit dieser Vermeidungskomponente relativ gute Schulnoten hatten. Im Gegensatz dazu lagen bei Schüler/-innen mit einer hohen FMp schlechtere Leistungsergebnisse und somit schlechtere Schulleistungen vor.

Aber auch das Vorliegen der beiden Komponenten FE und HM wird diskutiert. Dabei stammen die Überlegungen zum Motiv FE von Horner (1968). In ihren Untersuchungen fand sie heraus, dass Frauen häufiger Erfolg mit negativen Konsequenzen verbanden als Männer, also Erfolgsfurcht zeigten. Die Gegenkomponente dazu stellt das Motiv HM dar, nach welchem Misserfolg für Personen erstrebenswert sein kann. Die Überlegungen dazu gehen auf Swann (1983) zurück, in denen dieser argumentiert, dass Misserfolg für Personen erstrebenswert sein kann, damit sie ihr negatives Selbstkonzept aufrechterhalten können. In beiden Fällen stellen Leistungssituationen einen doppelten Appetenz-Aversions-Konflikt dar, da sowohl positive als auch negative Konsequenzen aus zwei möglichen Verhaltensalternativen folgen (Edelmann, 2000).

Hinzu kommen noch, neben den klassischen Leistungsmotiven, die beiden Komponenten Übermotivierte und Misserfolgsakzeptierende. Die Überlegungen zu diesen Typen gehen, wie bereits in Kapitel 1.1.2.3 beschrieben, auf Covington et al. zurück (Covington & Omelich, 1991; Covington & Roberts, 1994). Durch die Betrachtung der Komponenten der Leistungsmotivation als zwei unabhängige Dimensionen war es Covington und Roberts (1994) beispielsweise möglich, zahlreiche Unterschiede zwischen Übermotivierten und Misserfolgsakzeptierenden festzustellen. So sind mitunter übermotivierte Schüler/-innen dadurch gekennzeichnet, dass sie viel lernen und nach Erfolg streben, sich aber gleichzeitig vor Misserfolg fürchten und deshalb in Prüfungen oft Schwierigkeiten haben, ihr Wissen abzurufen. Für Misserfolgsakzeptierende hingegen

besitzen Leistungssituationen oft keinen Anreiz und sie hoffen weder auf Erfolg noch fürchten sie sich vor Misserfolg.

1.2 Intelligenz

Intelligenz kann als eines der am besten erforschten Merkmale der Psychologie gesehen werden (Rost, 2009). Dabei kann die Forschung der Intelligenz auf mehr als hundert Jahre Tätigkeit und auf eine kaum überschaubare Reichhaltigkeit an Befunden zurückblicken, welche in verschiedenen Disziplinen, auch außerhalb der Psychologie, intensiv und gleichzeitig kontrovers diskutiert werden. Hinzu kommt das Interesse der Öffentlichkeit an dem Konzept der Intelligenz und des Intelligenzquotienten (IQ), welches die Forschung in diesem Bereich weiter vorangetrieben hat (Rost, 2009).

Dieses Kapitel soll einen Überblick darüber geben, was man in der Psychologie unter Intelligenz versteht, wie diese gemessen wird, welche Strukturtheorien es dazu gibt und wie von Intelligenz auf andere Leistungen geschlossen werden kann.

1.2.1 Begriffsbestimmung von Intelligenz

Beim Durchlesen der gängigen Literatur zum Thema Intelligenz wird schnell klar, dass es trotz der intensiven Forschung in den letzten hundert Jahren weder eine einheitliche Auffassung darüber gibt, was Intelligenz ist, noch eine allgemein anerkannte Definition zu diesem Begriff vorliegt (Lohaus, Vierhaus & Maass, 2010). Das liegt wohl am ehesten an dem Umstand, dass es sich bei Intelligenz, anders als beim Alter oder Gewicht, nicht um ein einzelnes, direkt zu beobachtendes Merkmal handelt, sondern aus dem Verhalten erschlossen werden muss (Funke & Vaterrodt-Plünnecke, 2004). Darüber hinaus spielt auch die gewählte Perspektive der Psychologie, Intelligenz theoretisch zu erklären oder zu operationalisieren, für die Uneinigkeiten in der Literatur eine Rolle (Meer, 1998). So werden beispielsweise in der Informationsverarbeitung grundlegende Prozesse geistiger Leistung erforscht, wohingegen sich die Entwicklungspsychologie mit der Intelligenzentwicklung im Lebensverlauf beschäftigt (Funke & Vaterrodt-Plünnecke, 2004; Lohaus et al., 2010).

Ein Blick in ein Psychologisches Wörterbuch verrät, dass sich Intelligenz vom lateinischen Begriff „*intelligentia*“ ableitet und „vorwiegend die mit dem Verstand verbundenen geistigen Fähigkeiten in ihrer potentiellen und dynamischen Bedeutung“ (Häcker

& Stapf, 2009, S. 475) kennzeichnet. Die von diesem Begriff abgeleiteten, verschiedenen Definitionen beinhalten meist viele Gemeinsamkeiten. So verstand beispielsweise Wechsler unter Intelligenz „die zusammengesetzte oder globale Fähigkeit des Individuums, zweckvoll zu handeln, vernünftig zu denken und sich mit seiner Umgebung wirkungsvoll auseinanderzusetzen“ (1964, S. 13), wohingegen Sternberg im Jahre 1997 Intelligenz definierte als

diejenigen mentalen Aktivitäten, die sowohl für die Anpassung an äußere Gegebenheiten als auch für deren Veränderung und Auswahl notwendig sind... [I]ntelligenz reagiert nicht nur auf die Umwelt, sondern formt sie auch aktiv. Sie bietet Menschen die Möglichkeit, flexibel auf herausfordernde Situationen zu reagieren. (zitiert nach Bjorklund & Schneider, 2006, S. 770)

Beiden Definitionen gemeinsam ist die Annahme, dass es sich bei Intelligenz um die Fähigkeit handelt, angemessen auf die Umwelt zu reagieren, mithilfe des Denkens Aufgaben zu lösen, ohne dass hierfür bestimmte Erfahrungen notwendig sind und dass Intelligenz ein Ergebnis grundlegender mentaler Prozesse ist, welche sich in verschiedenen Zusammenhängen unterschiedlich stark manifestieren. Trotz vieler verschiedener verbaler Definitionen zum Intelligenzbegriff gibt es doch ein recht klares Verständnis darüber, welche spezifischen Fähigkeiten Intelligenz umfassen: Hierzu gehören unter anderem das Problemlösen sowie die Entscheidungs- und Integrationsprozesse (Meer, 1998). Zudem herrscht Einigkeit darüber, dass diese spezifischen Fähigkeiten mittels geeigneter Testverfahren über die Beobachtung von Leistung messbar sind. Einen kurzen Abriss über die Geschichte der Intelligenzmessung soll das nächste Kapitel geben.

1.2.2 Die Messung von Intelligenz

Eine Möglichkeit, sich dem Thema der Intelligenz zu nähern, ist die Psychometrie. Hierbei wird versucht, psychische Merkmale mithilfe von Tests zu messen. Dabei werden statistische Verfahren wie die Faktorenanalyse verwendet, um Intelligenztests zu analysieren. Aus diesen Ergebnissen wiederum werden dann mögliche Strukturen der Intelligenz abgeleitet (siehe Kapitel 1.2.3, Funke & Vaterrodt-Plünnecke, 2004).

Die Entwicklungen zur Messung der Intelligenz begannen vor mehr als hundert Jahren und gehen unter anderem auf die Arbeiten von Sir Francis Galton und James McKeen Cattell zurück (Häcker & Stapf, 2009). Galtons Interesse galt, als ein Verwandter von Darwin, besonders der Vererbung von Merkmalen und die Erfassung der Fähigkeiten von Menschen, wie beispielsweise die Intelligenz (Krohne & Hock, 2007).

Dabei entwickelte Galton eine Vielzahl von psychometrischen Verfahren, insbesondere Reaktionszeitmessungen, um die Unterschiede der Menschen in den allgemeinen kognitiven Fähigkeiten zu erfassen. Zudem konstruierte er Reizdiskriminationstests im visuellen, akustischen und kinästhetischen Bereich, da für Galton auch die Schärfe und die Unterscheidungsfähigkeit der Sinne ein Kennzeichen kognitiver Fähigkeiten war. Diese psychischen Variablen übertrug er anschließend in die Gauß'sche Normalverteilung und konnte so die überzufällige Häufigkeit von intellektuellen Hochleistungen in bestimmten Familien aufzeigen. Dieses war für Galton ein Beleg für die Vererbbarkeit von Intelligenz (Krohne & Hock, 2007). Cattell (1890) verfasste, unter anderem auf den Arbeiten von Galton aufbauend, seinen Aufsatz „Mental tests and measurements“. Die dort berichtete Testserie hatte das Ziel, Vorhersagen über die akademische Befähigung zu leisten, und zwar mithilfe von Unterscheidungsvermögenstests sowie Reaktionszeitmessungen. Cattell versuchte hierbei nicht nur körperliche und psychophysische Prozesse zu messen, sondern auch geistige Prozesse zu erfassen und verwies auch auf die Wichtigkeit der Standardisierung der Bedingungen (Häcker & Stapf, 2009). Die Arbeiten dieser beiden Forscher können als die psychometrischen Vorläufer der Intelligenzdiagnostik gesehen werden und ebneten den Weg für die weitere Intelligenztestforschung (Häcker & Stapf, 2009).

Der erste Intelligenztest wurde von Binet und seinem Mitarbeiter Simon (1903, 1905) entwickelt (zitiert nach Conrad, 1983). Dieses sogenannte Stufenleitermodell versuchte anhand eines einzigen Maßes Intelligenz bei Kindern zu beschreiben und zu differenzieren (Conrad, 1983), wobei aber nicht nur die individuelle Begabung, sondern auch das Lebensalter als Bestimmungsgröße für die Intelligenzleistung aufgefasst wurde: Verschiedene Altersgruppen erhielten entsprechend ihrem Alter unterschiedlich schwere Aufgaben zur Bearbeitung. Als Maß für die Intelligenz wurde das Intelligenzalter eingeführt, welches für die alterstypische Leistung steht (Häcker & Stapf, 2009). Dieses Intelligenzalter zeigt also anhand der gelösten Aufgaben, ob die Leistung des Kindes der alterstypischen Leistung entspricht, wobei das Lebensalter als Vergleichsmaßstab dient (Schweizer, 2006). Ein Nachteil des Intelligenzalters von Binet bestand darin, dass eine gleiche Abweichung zwischen Intelligenz- und Lebensalter, abhängig vom tatsächlichen Lebensalter, unterschiedliche Bedeutung hat, dieses aber unberücksichtigt bleibt (Schweizer, 2006). Zusätzlich kann dieses Verfahren nicht bei Erwachsenen angewandt werden, da hier laut dem Modell die Intelligenz mit zunehmendem Alter steigen müsste, was zu unrealistischen Ergebnissen führt.

Von dieser Entwicklung unabhängig führte Stern im Jahre 1912 den Intelligenzquotienten ein, der sich aus dem Quotienten von Intelligenz- und Lebensalter ergibt. Anschließend wird dieser Quotient standardmäßig mit 100 multipliziert, wodurch eine Übereinstimmung von Intelligenz- und Lebensalter auf den Wert 100 festgelegt wurde (Schweizer, 2006). Demnach entspricht ein IQ von 100 der Leistung des Altersdurchschnitts. Auch hier bestehen Nachteile für die Messung der Intelligenz bei Erwachsenen, da das Lebensalter steigt, aber das Intelligenzalter stagniert (Amelang & Bartussek, 1997). Im Jahre 1939 wurde der Intelligenzquotient als Abweichungskoeffizient von Wechsler neu definiert (Wechsler, 1964). Dieser Koeffizient gibt an, welche Position eine Person in der Populationsverteilung einnimmt und berücksichtigt, dass sich im Alter die geistigen Fähigkeiten verändern. Die einzelnen Populationen bestimmen sich dabei durch Altersgrenzen. Den IQ einer Person kann man folgendermaßen errechnen:

$$IQ = \frac{(X - M)}{SD} \times 15 + 100$$

X steht in dieser Gleichung für den Testscore, M für den Mittelwert der Normstichprobe und SD für die Standardabweichung der Normstichprobe. Der IQ-Wert ist damit populationsabhängig und kann in Bezug auf seine Population interpretiert werden. Auch heute noch wird diese Definition des Intelligenzquotienten von Wechsler in verschiedenen Intelligenztests verwendet.

So vielfältig wie die verbalen Definitionen sind auch die Anzahl und die Inhalte der heutigen Intelligenztests. Im deutschsprachigen Raum zählte Conrad im Jahr 1983 74 deutschsprachige Intelligenztests. Unter Intelligenztests werden nach Daumenlang (1990) diejenigen Tests verstanden, mithilfe derer sich die relative Ausprägung der intellektuellen Leistungsfähigkeit bestimmen lässt, wobei sie auf unterschiedliche Intelligenzdefinitionen und Intelligenztheorien gründen. Entsprechend der Intelligenztheorie, auf der ein Test basiert, misst er entweder das allgemeine Intelligenzniveau, wie beispielsweise der nonverbale Raven-Matrizen-Test (Raven, 2002), oder er bestimmt die Struktur der Intelligenz. Als Beispiele für Strukturintelligenztests können der Intelligenz-Struktur-Test 2000 (I-S-T 2000) von Amthauer et al. (1999, 2001) oder das Leistungsprüfsystem (LPS) von Horn (1962, 1983) genannt werden. Bei einer solchen Art von Tests wird auch von Strukturtests mit multiplen Faktoren gesprochen, da sie aus verschiedenen Einzelwerten bestehen (Amelang & Zielinski, 2002). Nach Brambring (1983) lässt sich diesen Strukturtests, welche zur Erfassung spezieller intellektueller Fähigkeiten dienen, neben den Einzelwerten auch durch Addition derselbigen das all-

gemeine Intelligenzniveau zuordnen. Zudem kann bei den heutigen Intelligenztests zwischen Individualtests, beispielsweise dem Hamburg-Wechsler-Intelligenztest, und Gruppentests, beispielsweise dem LPS, unterschieden werden (Conrad, 1983).

In einem Bericht, der den derzeitigen Stand der Intelligenzforschung widerspiegeln sollte, geben Neisser und Kollegen (1996) einen Überblick über die verschiedenen, bis dahin entwickelten Intelligenztests und Intelligenzmodelle. Dabei fanden sie unter anderem heraus, dass die Werte von verschiedenen Intelligenztests meist miteinander korrelieren, sodass Personen mit hohen Werten in einem Test meist auch hohe Werte in einem anderen Test aufweisen und umgekehrt (Neisser et al., 1996). Darüber hinaus haben sie sich in ihrem Bericht beispielsweise auch mit intellektuellen Unterschieden zwischen ethnischen und kulturellen Gruppen beschäftigt, aber auch mit solchen zwischen Männern und Frauen.

Dass im Rahmen dieser Arbeit nur ein kleiner Abriss über die Geschichte und Gegenwart der Messung der Intelligenz und der verschiedenen Tests gegeben werden kann, ist offensichtlich bei der Fülle der vorhandenen Tests, Modelle und der über hundertjährigen Forschungsdisziplin. Dennoch soll im Anschluss ein Überblick über die einflussreichsten und bedeutendsten Intelligenzmodelle gegeben werden.

1.2.3 Strukturtheorien der Intelligenz

Handelt es sich bei Intelligenz um eine einheitliche Fähigkeit oder besteht sie aus mehreren Komponenten? Welche Struktur der Intelligenz zugrunde liegt, wurde mit unterschiedlichen Methoden zu beantworten versucht. Diese Arbeit wird vorwiegend auf die Faktorenanalyse eingehen, welche die Grundlagen für die Strukturtheorien bildet. Das Verfahren der Faktorenanalyse wird kurz angerissen, bevor die für diese Studie wichtigsten Modelle der Theorien vorgestellt werden, da durch diese Analyse der Daten mehr über die Struktur der Intelligenz bekannt wurde (Roth, 1998).

Mit der Faktorenanalyse soll versucht werden, möglichst viele Variablen, die wechselseitig miteinander korrelieren, zu möglichst wenigen Dimensionen bzw. Faktoren zusammenzufassen. Dabei stellen die Dimensionen die Komponenten intelligenten Verhaltens des Menschen dar (Jäger, 1973). Der Faktor, der durch die Faktorenanalyse entsteht, sollte inhaltlich mit den Variablen, die zu ihm gehören und mit ihm korrelieren, einen nach Möglichkeit großen Anteil des Gemeinsamen umfassen. Dabei stellt die Variablenzusammenfassung in Faktoren, insbesondere bei den großen Mengen an Variab-

len und der daraus resultierenden Vielzahl von Beziehungen bei Intelligenztests, eine Vereinfachung der Darstellung von Korrelationen dar. Es zeigte sich, dass in den Modellen der Strukturtheorien der Intelligenz die Vorgehensweise der Faktorenanalyse die Resultate determiniert. Mit den ersten beiden Modellen, dem Zweifaktorenmodell von Spearman (1904) und dem Primärfaktorenmodell von Thurstone (1938), einem Modell mit mehreren gemeinsamen Faktoren, standen sich entgegengerichtete Modelle gegenüber. Auf diese beiden und auf weitere, für diese Arbeit wichtige Modelle, soll im Folgenden eingegangen werden.

Zusammenfassend lässt sich festhalten, dass das Ziel der Faktorenanalyse also darin besteht, eine nach Möglichkeit große Anzahl unterschiedlicher kognitiver Aufgaben auf die Faktorenstruktur zu untersuchen und diese in möglichst wenigen Dimensionen zusammenzufassen.

1.2.3.1 Die Zwei-Faktoren-Theorie von Spearman (1904)

Das erste Strukturmodell der Intelligenz, das Zwei-Faktoren-Modell, wurde 1904 von Spearman formuliert. In diesem definierte er Intelligenz als eine alle kognitiven Fähigkeiten beeinflussende Eigenschaft. Um die Faktoren der Intelligenz bestimmen zu können, benutzte er als einer der ersten die Faktorenanalyse. In seinen Analysen bemerkte Spearman, dass alle Intelligenztests bzw. Aufgaben eines Intelligenztests, Schulleistungen und Einschätzungen der Lehrer einer Versuchsperson positiv miteinander korreliert waren. Aus diesem Grund kam er zu dem Schluss, dass jedes Intelligenzmaß auf zwei Faktoren beruhen müsse. Den ersten Faktor nannte er Generalfaktor (kurz *g*) bzw. gemeinsamen Faktor. Deshalb findet sich in der Literatur auch teilweise die Bezeichnung „Generalfaktormodell“ für das Intelligenzmodell von Spearman (Conrad, 1983). Dieser *g*-Faktor besagt nach Spearman, dass jede Person über ein bestimmtes Maß an allgemeiner Intelligenz verfügt, welches in kognitiven Aufgaben sowohl die Denk- als auch die Lernfähigkeit bestimmt. Die Unterschiede einer Person in den verschiedenen Testaufgaben bzw. Testverfahren, also Faktorladungen, welche nicht auf *g* luden, erklärte er durch die spezifischen Faktoren für den jeweiligen Test, die *s*-Faktoren. Somit geht nach Spearman die Varianz in Intelligenztests teils auf *g* zurück und ist teils spezifisch. Die Formulierung des *g*-Faktors und die damit verbundene Definition der allgemeinen Intelligenz sind bis heute ein zentrales Merkmal der Theorie von Spearman, und alle späteren hierarchischen Theorien der Intelligenz gehen darauf zurück (Roth, 1998). Ein

Intelligenztest, welcher auf der Zwei-Faktoren-Theorie von Spearman beruht, ist beispielsweise der Raven-Matrizen-Test (Conrad, 1983).

1.2.3.2 Das Primärfaktorenmodell von Thurstone (1938)

Thurstone untersuchte in Korrelationsmatrizen mithilfe der von ihm entwickelten multiplen Faktorenanalyse die substantielle Varianz, die nach Extraktion eines g -Faktors noch immer vorhanden war und nicht erklärt werden konnte (Funke & Vaterrodt-Plünnecke, 2004). Aufgrund dessen ging Thurstone (1938) in seinem Primärfaktorenmodell, im Gegensatz zu Spearman, nicht von einem einzigen g -Faktor aus, sondern von mehreren nebeneinander stehenden, voneinander unabhängigen und gleichberechtigten Faktoren, den sogenannten Primärfaktoren. Dabei spiegeln diese Faktoren grundlegende Komponenten der Intelligenz wider und sind beim Bearbeiten verschiedenster kognitiver Aufgaben in unterschiedlicher Gewichtung beteiligt. Die von Thurstone postulierten sieben Primärfaktoren – Rechenfähigkeit, Sprachverständnis, Wortflüssigkeit, Kurzzeitgedächtnis und Merkfähigkeit, Erkennen von Regelmäßigkeit und schlussfolgerndes Denken, Wahrnehmungsgeschwindigkeit und räumliches Vorstellungsvermögen – konnten in mehreren faktorenanalytischen Studien und von unabhängigen Forschern bestätigt werden (Funke & Vaterrodt-Plünnecke, 2004). Das von Thurstone vorgeschlagene Modell findet sich unter anderem im LPS von Horn (1962, 1983) wieder.

Besteht mit der Formulierung der Primärfaktoren von Thurstone ein unüberwindbarer Gegensatz zum Generalfaktorenmodell von Spearman? Mit dieser Frage haben sich in der Vergangenheit viele Intelligenzforscher beschäftigt und festgestellt, dass diese beiden auf den ersten Blick unterschiedlichen Theorien nicht so verschieden sind und zudem nicht im direkten Widerspruch stehen. So weisen Funke und Vaterrodt-Plünnecke (2004) auf die unterschiedliche Stichprobenauswahl der beiden hin: Im Vergleich zu Spearman hat Thurstone in seinen Untersuchungen mit homogeneren Stichproben von Studierenden gearbeitet, die aufgrund der Varianzeinschränkung zu mehr Faktoren bzw. spezifischen Merkmalen geführt haben. Aus diesem Blickwinkel heraus erscheinen die beiden Modellvorstellungen nicht in direktem Widerspruch zueinander zu stehen.

1.2.3.3 Fluide und kristalline Intelligenz nach Cattell (1963, 1971)

Eine Weiterentwicklung der hierarchischen Intelligenzmodelle wurde von Cattell (1963, 1971) vorgenommen. Den Kern seines Modells bilden zwei voneinander unabhängige Faktoren der Intelligenz, welche er fluide Intelligenz (g_f) und kristalline Intelligenz (g_c) nannte. Dabei steht das g für die Generalität dieser Faktoren. Unter fluider Intelligenz versteht Cattell (1963, 1971) die Fähigkeit, ohne Vorwissen neuartige Probleme und Situationen erfolgreich zu bewältigen. Diese Komponente des geistigen Leistungsvermögens basiert auf der vererbten Funktionstüchtigkeit der hirnelementarbiologischen Prozesse. Im Gegensatz dazu bezeichnet nach Cattell (1963, 1971) die kristalline Intelligenz die umweltbedingte Komponente, welche durch Lernvorgänge bestimmt wird. Diese wird also durch Lernerfahrungen und Faktenwissen erworben und ist für die Verarbeitung von vertrauten Informationen, aber auch bei der Wissensanwendung relevant (Cattell, 1963, 1971). Für die Unterscheidung in zwei Faktoren sprechen ihre unterschiedlichen Entwicklungsverläufe. So wächst die kristalline Intelligenz von der Kindheit bis ins Alter kontinuierlich an, wohingegen die fluide Intelligenz im frühen Erwachsenenalter ihren Höhepunkt erreicht und mit zunehmendem Alter wieder abnimmt (Cattell, 1963, 1971). Ein Intelligenztest, welcher diesem Modell folgt, ist der Grundintelligenztest (CFT), beispielsweise der CFT 3 von Cattell (1971). Dieser Test findet häufig bei schulischen Fragestellungen Anwendung (Funke & Vaterrodt-Plünnecke, 2004).

1.2.3.4 Das Drei-Ebenen-Modell nach Carroll (1993)

Ein weiteres hierarchisches Modell, welches die Intelligenz mithilfe dreier Ebenen beschreibt, wurde von Carroll (1993) generiert. Um dieses Modell entwickeln zu können, verwendete er Datensätze zur Intelligenzmessung aus der Vergangenheit und fasste diese zu Korrelationsmatrizen zusammen. Anschließend führte er an diesen 461 Datensätzen Faktorenanalysen und Vergleiche durch. Auf diese Weise konnte Carroll eine große Bandbreite kognitiver Fähigkeiten bei Menschen identifizieren und daraus entstand sein hierarchisches Intelligenzmodell, bestehend aus drei Ebenen. Auf der obersten Ebene, auch Stratum III genannt, befindet sich die allgemeine Intelligenz, welche sich auf alle Intelligenzfähigkeiten auswirkt. Sie entspricht dem von Spearman postulierten g -Faktor und beeinflusst die kognitiven Leistungen auf der nächsten Ebene bzw. dem Stratum II,

zu welcher unter anderem die fluide und kristalline Intelligenz nach Cattell (1963) sowie sechs weitere Komponenten der Intelligenz gehören (siehe Abbildung 1.2).

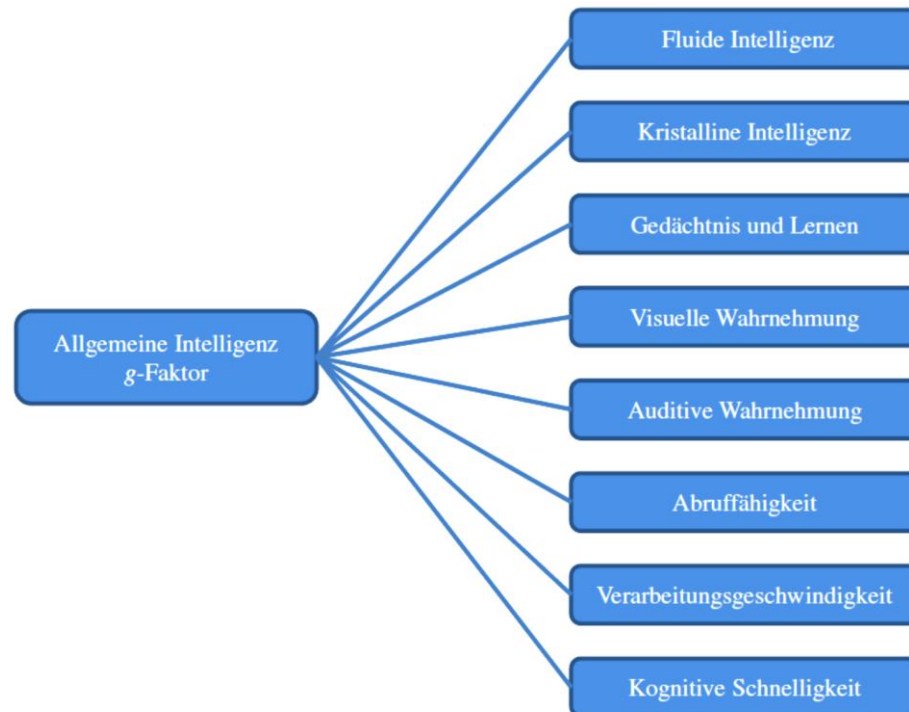


Abbildung 1.2

Stratum III und II des Drei-Ebenen-Modells von Carroll (1993, S. 626)

Zur untersten Ebene, auch Stratum I genannt, gehört eine Vielzahl von spezifischen Faktoren, welche mit den von Spearman postulierten *s*-Faktoren Ähnlichkeit aufweisen und die sowohl durch die allgemeine Intelligenz als auch durch die Intelligenzkomponenten der mittleren Ebene beeinflusst werden. So zählen zu diesen spezialisierten Faktoren beispielsweise Fähigkeiten wie das Schlussfolgern oder die Geräuschunterscheidung (Carroll, 1996). Dabei repräsentieren diese Faktoren Effekte des Lernens oder kognitive Fähigkeiten und können durch verschiedene Testungen direkt erhoben werden. Darüber hinaus gibt Carroll (1993) auch Aufgabenbeschreibungen, die notwendig sind, um die beschriebenen Faktoren zu erheben. Somit entsprechen diese Faktoren realen Phänomenen, welche die kognitiven Fähigkeiten steuern, unabhängig zu welcher Ebene sie gehören. Diese Faktoren können bei einer Testbearbeitung konkret festgestellt werden. Zusammenfassend soll das Modell von Carroll (1993) einen Rahmen bilden, in dem sowohl Variablen der Psychometrie als auch der Informationsverarbeitung inte-

griert sind und für Interpretationen genutzt werden können. Dieses Drei-Ebenen-Modell bietet wiederum den theoretischen Rahmen für das von Kreuzpointner (2010) weiterentwickelte LPS-neu, indem jenes ebenfalls eine Unterscheidung in Stratum I, II und III vornimmt. Dabei ist auch beim LPS-neu nach Kreuzpointner (2010) die allgemeine Intelligenz auf der obersten Ebene angesiedelt und hat Auswirkungen auf alle Intelligenzfähigkeiten. Im Gegensatz zu Carroll (1993) legt Kreuzpointner (2010) in Stratum II eine Unterteilung der Intelligenz in nur vier statt in acht Komponenten zugrunde: die kristalline und fluide Intelligenz, die visuelle Wahrnehmung sowie die kognitive Schnelligkeit. Diese vier Intelligenzkomponenten werden durch die elf Subtests des LPS-neu erfasst, welche die unterste Ebene darstellen. Das LPS-neu wird, da es in dieser Studie als Leistungstest verwendet wird, im Kapitel 3.4.6 noch näher erläutert werden.

Eine Weiterentwicklung sowohl der Drei-Ebenen-Theorie von Carroll als auch der Theorie der fluiden und kristallinen Intelligenz von Cattell findet sich in der Cattell-Horn-Carroll-Theorie (CHC-Theorie), welche die beiden erstgenannten Theorien integriert (McGrew, 2005). Dabei basiert diese Theorie auf einer umfangreichen Faktorenanalyse und wurde zudem bei vielen neu entwickelten Verfahren verwendet. So beziehen sich beispielsweise der Hamburg-Wechsler-Intelligenztest für Kinder IV (HAWIK-IV) oder die Kaufman Assessment Battery for Children (K-ABC-II) auf dieses Modell (Daseking, Janke, & Petermann, 2006).

Die vorgestellten Intelligenzstrukturmodelle zeigen nur einen kleinen Ausschnitt der bis heute entwickelten Intelligenztheorien. Trotz dieser Fülle an verschiedenen Theorien und der fast hundert Jahre dauernden Forschung konnte bis heute noch kein einheitliches, allgemein anerkanntes Intelligenzstrukturmodell entwickelt werden. Dennoch herrscht Einigkeit darüber, dass es sich bei Intelligenz um ein komplexes Phänomen handelt und dass für die Entwicklung von Strukturmodellen objektive Messungen und Faktorenanalysen eine bedeutende Rolle spielen (Roth, 1998). Nur so kann ein Erkenntnisfortschritt erzielt werden, der zu neuen Theorien führt, die wiederum die Erfassung des Intelligenzkonstrukts möglich machen (Roth, 1998).

1.2.4 Intelligenz und Schulleistung

Im Laufe der langen Forschungstradition wurde die Bedeutung der Intelligenz auch für verschiedene Lebensbereiche untersucht. Der für diese Arbeit interessante Zusammen-

hang zwischen Intelligenz und Schulerfolg wird im Folgenden anhand einer kleinen Auswahl empirischer Untersuchungen und Ergebnissen vorgestellt.

Rost (2009) zufolge kann die allgemeine Intelligenz bzw. g als bester Einzelprädiktor für die Schulleistung gesehen werden. Dabei stützt er sich sowohl auf die Feststellung von McNemar, Präsident der American Psychological Association, der dieses nach Durchsicht von über 4000 Studien bereits 1964 betonte, als auch auf die Übersicht von Kühn (1987), welcher 24 deutschsprachige Intelligenztests im Hinblick auf die statistische Vorhersage des Schulerfolgs untersuchte. Kühn (1987) kam dabei unter anderem zu dem Ergebnis, dass sich Zensuren in den Hauptfächern besser vorhersagen lassen als Zensuren in den Nebenfächern. Auch Funke und Vaterrodt-Plünnecke (2004) weisen darauf hin, dass die Korrelation sowohl bei Mathematik- als auch bei Deutschnoten mit Intelligenztestwerten am höchsten ist. Zudem verweist Rost (2009) auf eine Untersuchung von Leeson, Ciarrochi und Heaven aus dem Jahre 2008, in der sowohl Zensuren, verbale bzw. numerische Leistungsfähigkeit als auch Persönlichkeitsvariablen erhoben wurden. Es zeigte sich, dass deutlich höhere Korrelationen zwischen den Intelligenzdomänen und Zensuren vorlagen als zwischen Persönlichkeitsvariablen und Schulleistung (Rost, 2009).

Vergleiche von Intelligenzergebnissen aus standardisierten Tests mit den Kriterien für den Schulerfolg weisen dementsprechend hohe Übereinstimmungen auf. So konnten etwa Neisser und Mitarbeiter (1996) in ihrem Bericht im Mittel eine Korrelation von $r = .50$ zwischen Intelligenz und Schulleistung in der Grundschulzeit aufweisen. In einer Untersuchung von Wechsler (1964) korrelierte die schulische Ausbildung, gemessen in Jahren, mit den Intelligenztestergebnissen für eine Erwachsenenstichprobe bei $r = .66$. Auch bei Mayes und Kollegen (2009) korrelierte die Schulleistung positiv mit den Intelligenztestergebnissen von Elementarschüler/-innen. Dieser Zusammenhang ist dabei nach Jensen (1998) in der Grundschulzeit am höchsten, sinkt dann mit steigendem Ausbildungslevel und ist auf dem Universitätsniveau am geringsten. Als Erklärung für diesen Rückgang finden sich in der Literatur diverse Gründe. So macht Rost (2009) zum Beispiel auf den Umstand aufmerksam, dass ein Sinken der Intelligenzvarianz durch die Aussonderung leistungsschwacher Schüler/-innen zwangsläufig zur Abnahme der Korrelation führt. Helmke und Weinert (1997) vermuten, dass das Vorwissen einer Person mit steigendem Alter und Ausbildungsniveau stetig relevanter wird und erklären den Rückgang der Korrelation auf diese Weise. Dementsprechend müsste der Schulerfolg in den höheren Klassen bzw.

am Ende einer Schullaufbahn vorrangig vom Wissen abhängen, welches über die Schuljahre hinweg aufgebaut wurde. Weitere Gründe könnten sich aber auch in der geringeren Leistungsbreite auf Universitätsniveau oder in Persönlichkeitseinflüssen finden, die über eine so langfristige Dauer gesehen nicht genau vorhergesagt werden können (Amelang, 1976).

Festzuhalten bleibt, dass es einen bedeutsamen Zusammenhang zwischen Intelligenz und Schulerfolg gibt und dass dieser zu einem der höchsten in der psychologischen Diagnostik zählt. Dennoch dürfen andere Faktoren, die unabhängig von der Intelligenz ebenfalls einen Einfluss auf die Schulleistung haben können, nicht außer Acht gelassen werden. So können etwa die Erwartungen der Eltern und Lehrer, der sozioökonomische Status der Familie, die implizite Intelligenztheorie der Schüler/-innen, die kulturellen Faktoren und die intrinsische Motivation der Schüler/-innen, aber auch Aufmerksamkeit, Anstrengung, Lernstrategien sowie Konzentration, einen Einfluss auf die schulischen Leistungen haben (vgl. hierzu Funke & Vaterrodt-Plünnecke, 2004; Helmke & Weinert, 1997).

1.3 Testmotivation

„Der Patient war anscheinend motiviert..., doch die Testergebnisse sind ungültig“ (Green, 2004, S. 303). Viele Psychologen, die in Deutschland in der klinischen Leistungsdiagnostik tätig sind, erfassen bzw. messen die Testmotivation meist nicht routinemäßig. Doch weshalb ist die Erfassung der Testmotivation für die Psychologie so wichtig? Wird sie nicht gemessen, besteht die Gefahr, Testwerte als valide zu betrachten, obwohl sie es nicht sind. So konnten zum Beispiel Green und Kollegen (2001) in einer Untersuchung zeigen, dass 50 Prozent der Varianz der Testergebnisse durch die Leistungsmotivation erklärt werden können. Dabei stellte die Testmotivation mit ihrem Einfluss auf andere Testergebnisse die wichtigste Einzelvariable dar. Aber nicht nur in Einzel-, sondern auch in Gruppenuntersuchungen sollte die Testmotivation gemessen werden, da die sonst erhaltenen Daten zu ungültigen Testergebnissen und somit zu falschen Schlussfolgerungen führen können (Merten & Brockhaus, 2004). Wird also ohne Kontrolle dieser wichtigen Variable geforscht, kann es zu irreführenden Ergebnissen kommen. Dieses wiederum kann die psychologische Theorienbildung beeinflussen. Dabei betrifft die Testmotivation nicht nur die Neuropsychologie, sondern auch andere Bereiche der Psychologie, wie etwa die angewandte Psychologie oder die grundlagen-

orientierte Forschung (Green, 2004). Aus diesen Gründen beschäftigt sich die vorliegende Arbeit mit dem Thema der Testmotivation, wobei zunächst eine Begriffsbestimmung folgen wird.

1.3.1 Begriffsbestimmung von Testmotivation

Auch für den Begriff der Testmotivation gibt es, genauso wie für die Leistungsmotivation und die Motivation allgemein, verschiedene Definitionen und Begriffsklärungen. So verstehen Wise und DeMars (2005) unter Testmotivation das Engagement und den Energieaufwand, den eine Person in Bezug auf ein Ziel aufbringt, um das bestmögliche Ergebnis in einem Test zu erreichen. Sie weisen aber darauf hin, dass sich in der Forschung eine breitere Definition, die zudem Motivationsmessungen wie Anstrengung, Interesse, Nützlichkeit, Wert und Wichtigkeit berücksichtigt, durchgesetzt hat (Wise & DeMars, 2005). Eine frühere und einfachere Definition stammt von Salgado, Remeseiro und Iglesias (1996). Für sie ist Testmotivation die positive oder negative Einstellung, einen Test zu beantworten. Eklöf (2008) fügt ihrer Definition von Testmotivation noch hinzu, dass sie situationsspezifisch ist. Demnach kann laut Eklöf (2008) die Leistungsmotivation begrifflich geordnet und auf einer situationsspezifischen Stufe gemessen werden. Die situationsspezifische Motivation, also die Testmotivation, ist die Motivation, in einer gegebenen Situation oder in einem gegebenen Test erfolgreich zu sein. Dabei stellt die Testmotivation eine situationsspezifische Komponente der Leistungsmotivation dar, die nur in bestimmten Situationen bzw. bei bestimmten Tests angesprochen wird (Eklöf, 2008). Eklöf (2008) begründet ihre Definition auf der Aussage von Messick aus dem Jahr 1988. Messick bemerkte, dass eine schlechte Testleistung nicht nur aufgrund des Testinhalts und der Fähigkeiten der Studenten/-innen zu interpretieren sei, sondern auch ein Ausdruck von mangelnder Motivation sein könnte. Wenn sich verschiedene Gruppen von Personen in ihrem Motivationsniveau systematisch unterscheiden und wenn weniger motivierte Personen deshalb benachteiligt sind, weil ihr Ergebnis unter dem ihres aktuellen Potenzials bleibt, dann könnte Testmotivation eine mögliche Quelle von Verzerrungen sein und wäre wichtig für die Interpretation der Ergebnisse (Eklöf, 2008). Somit könnte die Testmotivation ein entscheidender Aspekt für die Validität und die Zuverlässigkeit der Testergebnisse (Eklöf, 2008) sein.

Auch Salgado et al. (1996) weisen darauf hin, dass diese Motivation einen signifikanten Effekt auf die Größe der Validitätskoeffizienten haben könnte, aber hierzu bis-

lang zu wenig Forschung betrieben worden ist. Dass die Forschung der Testmotivation aus mehreren Gründen wichtig ist, unter anderem wegen der Einflussnahme auf die aktuelle Testvalidität, befürworten auch Sanchez, Truxillo und Bauer (2000). Zudem ist nach Baumert und Demmrich (2001) das Wissen darüber, wie eine Person den Test, den sie gerade bearbeitet, wahrnimmt und wie die Motivation, in diesem Test erfolgreich zu sein, die Person beeinflusst, relativ rar und sollte aus diesen Gründen ebenfalls weiter erforscht werden.

Ob die Testmotivation tatsächlich einen Einfluss auf die Testergebnisse und die Validitätskoeffizienten oder einen Einfluss auf andere Komponenten in einem Test nimmt, wird im Kapitel 1.3.3 – Empirische Befunde zur Testmotivation – behandelt. Zunächst wird ein mögliches Modell der Testmotivation vorgestellt.

1.3.2 Modell der Testmotivation

Als Basis für das Modell der Testmotivation dient speziell die von Eccles und Kollegen (1983) erstellte Erwartungs-Wert-Theorie der Leistungsmotivation. Dabei haben Eccles et al. (1983) eine Erwartungswerttheorie entwickelt, die auf dem Erwartungs-Wert-Modell von Atkinson (1964/1975) aufbaut, indem sie die Fähigkeit, die Ausdauer und die Wahl des Leistungsverhaltens direkt auf eine Erwartungs- und eine Wertkomponente zurückführen. Die Theorie unterscheidet sich jedoch auch von Atkinsons Modell, da Eccles et al. (1983) davon ausgehen, dass die Erwartungs- und Anreizkomponenten mit einem weiteren Feld von psychologischen und sozialen/kulturellen Determinanten verbunden sind. Zudem wird angenommen, dass Erwartung und Anreiz in positiver Beziehung stehen und nicht invers, wie Atkinson vorgeschlagen hat (Eccles & Wigfield, 2002). Der erste Entwurf dieser Theorie kann der Abbildung 1.3 entnommen werden.

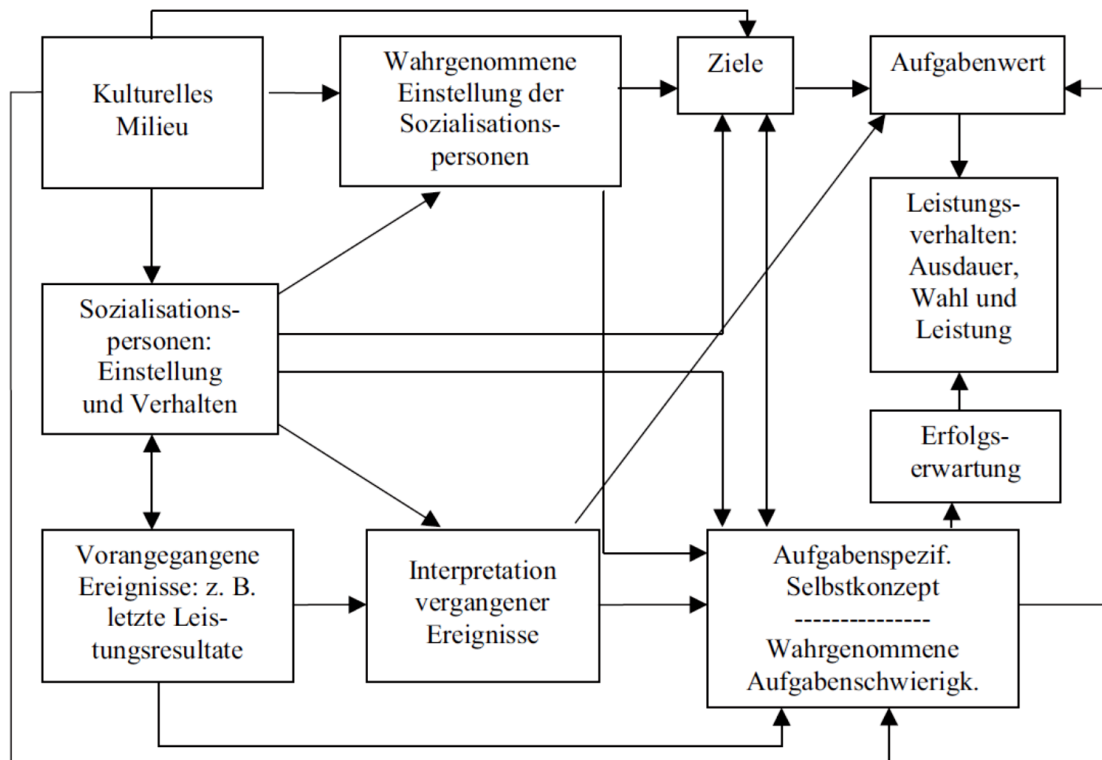


Abbildung 1.3

Erwartungs-Wert-Modell nach Eccles et al. (1983) (zitiert nach Pohlmann, Möller & Streblow, 2005, S. 128)

Wigfield und Eccles (2000) nehmen an, dass die Erwartung und die Wertigkeit sowohl die Leistungswahl als auch die Fähigkeit, die Anstrengung und die Ausdauer des Leistungsverhaltens direkt beeinflussen. Zudem gehen sie davon aus, dass die Erwartung und die Wertigkeit durch den aufgabenspezifischen Glauben, wie den Fähigkeitsglauben, die wahrgenommene Schwierigkeit der Aufgabe und die individuellen Ziele, wie Selbstschemata und affektive Erinnerungen, beeinflusst werden. Diese sozial kognitiven Variablen werden dabei durch individuelle Wahrnehmungen der eigenen Erfahrungen und von einer Reihe sozialer Einflüsse beeinflusst (Eccles & Wigfield, 2002).

Das Modell kann, wie Studien unter anderem von Wolf und Smith (1995) bzw. von Wolf, Smith und Birnbaum (1995) (siehe dazu Kapitel 1.3.2.1) belegen, auch für das Konstrukt der situationsspezifischen Motivation, der Testmotivation, verwendet werden. Angewandt auf die Testmotivation besagt dieses Modell, dass die situationsspezifische Motivation sowohl von der Erwartung einer Person, wie gut sie in der bevorstehenden Aufgabe abschneiden wird, als auch von dem Aufgabenwert, also dem subjektiven Wert dieser Aufgabe für die Person, beeinflusst wird (Eklöf, 2006; Frey, Hartig & Moosbrugger, 2009; Jacobs, 2009). Infolgedessen beeinflussen diese beiden Kompo-

nenten das Leistungsverhalten einer Person direkt, welches sich in Ausdauer, Wahl und Leistung bzw. Anstrengung widerspiegelt (Pohlmann et al, 2005). Liegt also eine entsprechende Testmotivation vor, so zeigt die Person eine höhere Ausdauer und Anstrengung für die Bearbeitung der Aufgaben und sucht sich zudem, bei freier Aufgabenwahl, Aufgaben, die sie herausfordern.

Gemäß der Modellannahmen wird die Erfolgserwartung sowohl vom Begabungselbstkonzept als auch von der wahrgenommenen Aufgabenschwierigkeit direkt bestimmt (Pohlmann et al., 2005). Unter dem Fähigkeitsselbstkonzept verstehen Eccles und Kollegen (1983) die Kompetenzeinschätzung einer Person in einer gegebenen Aktivität. Die beiden Konstrukte unterscheiden sich also dahingehend, dass die Erfolgserwartung auf zukünftige und das Fähigkeitsselbstkonzept auf gegenwärtige Aufgaben fokussiert ist (Wigfield & Eccles, 2000). In Leistungssituationen hingegen konnte gezeigt werden, dass sich diese Konstrukte empirisch nicht voneinander unterscheiden (Eccles & Wigfield, 2000). Die aufgabenbezogenen Kognitionen, also das aufgabenspezifische Selbstkonzept und die wahrgenommene Aufgabenschwierigkeit, auf welche individuelle Ziele, bisherige Erfolgs- und Misserfolgserfahrungen sowie das kulturelle Milieu Einfluss nehmen, werden wiederum durch die Interpretationen vergangener Ereignisse und die Einstellung bzw. wahrgenommene Einstellung der Bezugspersonen beeinflusst (Pohlmann et al, 2005). Fühlt sich die Person demnach in einer gegebenen Aufgabe kompetent und wird die subjektive Schwierigkeit einer Aufgabe als niedrig bzw. lösbar eingeschätzt, so steigt die subjektive Erfolgserwartung und die Person glaubt daran, diese Aufgabe gut zu meistern. Dabei wird sowohl die Kompetenzeinschätzung für eine Aufgabe als auch die wahrgenommene Aufgabenschwierigkeit vor allem von Erfahrungen in bisherigen Leistungssituationen und eigenen zukünftigen Bestrebungen beeinflusst. Die subjektive Aufgabenschwierigkeit wird zudem meist noch als abhängig vom Leistungsergebnis der anderen Personen gesehen, weshalb sie von sozialen Normen abhängig ist (Neugebauer, 1997).

Zum Aufgabenwert gehören nach Eccles und Kollegen (1983) vier Komponenten: die Wichtigkeitskomponente, der intrinsische Wert, die Nützlichkeitskomponente und die Kosten. Als Leistungswert bzw. Wichtigkeitskomponente definieren sie, wie wichtig es für eine Person sei, eine Aufgabe erfolgreich zu lösen. Dabei wird die Relevanz der Aufgabe oftmals für die Bestätigung hervortretender Aspekte des Selbstschemas verbunden (Pohlmann et al., 2005). So ist der Wert einer Aufgabe höher, wenn diese mit bedeutenden Selbstkonzeptaspekten, wie etwa der Kompetenz in verschiedenen

Domänen, als kongruent wahrgenommen wird (Eccles et al., 1983; Eccles & Wigfield, 2002). Unter dem intrinsischen Wert wird die Freude einer Person bei der Ausübung einer Aufgabe oder das subjektive Interesse der Person mit dem Thema verstanden (Eccles et al., 1983; Eccles & Wigfield, 2002). Die Nützlichkeitskomponente nimmt Bezug auf das Ausmaß, inwieweit eine Aufgabe für gegenwärtige oder zukünftige Ziele relevant ist. Dabei kann eine Aufgabe einen positiven Wert für das Erreichen erwünschter Zielzustände haben, obwohl die Person gegenwärtig nicht daran interessiert ist. Aus diesen Gründen erfasst diese Komponente eher extrinsische Gründe für die Ausübung einer Tätigkeit (Eccles et al., 1983; Eccles & Wigfield, 2002). Zudem konnten sie noch eine weitere Komponente identifizieren: die Kosten. Dabei sind Kosten unter anderem durch negative Aspekte wie die Leistungsangst und die Angst sowohl vor Misserfolg als auch vor Erfolg gekennzeichnet. Hinzu kommen die Aspekte, dass man sich, um erfolgreich zu sein, anstrengen und dabei zudem andere Tätigkeiten ausschließen muss, die man dann nicht mehr machen kann, wenn man sich bereits für eine Aufgabe entschieden hat (Eccles & Wigfield, 2002). Diese vier Komponenten sind wiederum abhängig von den individuellen Zielen einer Person, dem aufgabenspezifischen Selbstkonzept sowie der Interpretation vorausgegangener Ereignisse. Dabei werden die Ziele durch die Bezugspersonen und das kulturelle Milieu beeinflusst (Pohlmann et al., 2005). Demnach ist der subjektive Wert einer Aufgabe für eine Person umso höher, je wichtiger ihr diese ist, je mehr Freude sie bei der Ausübung jener Aufgabe hat, je höher der wahrgenommene Nutzen dieser Aufgabe ist und je weniger Kosten ihr, im Sinne von Leistungsangst bzw. Misserfolg, dabei entstehen.

Die postulierten Zusammenhänge im Erwartungs-Wert-Modell konnten inzwischen für akademische Zusammenhänge hinreichend bestätigt werden (Pohlmann et al., 2005). So konnten Eccles und Kollegen unter anderem zeigen, dass das Fähigkeitsselbstkonzept und die Erfolgserwartung die Leistung, etwa in Mathematik, beeinflusst, wohingegen der Aufgabenwert zum Beispiel die Kurswahl bestimmt (Eccles & Wigfield, 2002).

Zusammenfassend besagt die Theorie, dass eine bessere Fähigkeit hinsichtlich des domainspezifischen Inhalts eines Testes zu einem höheren Fähigkeitsglauben, zu einer höheren Erfolgserwartung und konsequenterweise zu einer höheren Motivation in diesem Test führen sollte (Eccles et al., 1983; Eccles & Wigfield, 2002). Des Weiteren besagt das Modell: Wenn alle anderen Faktoren konstant gehalten werden, bringen Personen mit höheren Fähigkeiten auch höhere Anstrengungen auf, um die Aufgaben in einem Test zu lösen (Segal, 2006). Demnach würde, wenn sich Personen mit denselben

Fähigkeiten in ihrer Testmotivation unterscheiden, dieses durch die unterschiedliche Anstrengungsbereitschaft in der Bearbeitung eines Tests sichtbar werden. Folglich wäre ein Test, für den alle das nötige Wissen haben, um ihn korrekt zu beantworten, geeignet, um mehr über individuelle Unterschiede in der Testmotivation zu erfahren (Segal, 2006).

1.3.3 Empirische Befunde zur Testmotivation

Aus welchen Gründen die Erfassung der Testmotivation wichtig ist, wurde bereits in den vorherigen Unterkapiteln zur Testmotivation erwähnt. Aber lassen sich dafür auch empirische Befunde finden? Dieses Kapitel wird einen Abriss über die bisherigen Untersuchungen und Ergebnisse zur Testmotivation, einer im Vergleich zur Leistungsmotivation noch sehr jungen Disziplin, liefern. Dabei wird zunächst der Einfluss der Testmotivation auf die Testleistungen in Tests mit und ohne Konsequenzen, geklärt und anschließend auf die Auswirkungen der Testmotivation speziell in Leistungstests eingegangen.

1.3.3.1 Der Einfluss von Testmotivation auf die Testleistungen

Um den Einfluss der Testmotivation auf die Testleistungen erfassen zu können, sollte die Erhebung, da die aktuell in einer Leistungssituation angeregte Motivation in Abhängigkeit von der jeweiligen Testsituation variiert, nach verschiedenen Studien (Frey et al., 2009; Arvey et al., 1990; Chan et al., 1998) möglichst kurz vor, während oder – unter anderem nach Arvey et al. (1990), Eklöf (2006) und Sanchez et al. (2000) – kurz nach dem Handlungsvollzug stattfinden. Dadurch soll versucht werden, das motivationale Resultat, welches sich aus dem Zusammenwirken von Person- und Situationsmerkmalen ergibt, zu quantifizieren (Frey et al., 2009). Zudem kann so der Einfluss der Testmotivation auf die Testleistung vor einem Test erfasst werden und anschließend außerdem der Einfluss der wahrgenommenen Leistung in einem Test auf die Motivation untersucht werden (Sanchez et al., 2000).

Die Messung der Testmotivation fand zuerst im Kontext der Personalauswahl statt. Dabei haben Arvey und Kollegen (1990) ein diagnostisches Verfahren entwickelt, das Test Attitude Survey (TAS), welches eine Motivationssubskala (TAS-M) enthält. In diesem Verfahren werden die Testteilnehmer direkt nach einem Test über die Testan-

strengung und -einfachheit sowie über die Motivation im Test und über das Gefallen des Tests befragt. Mithilfe dieses Verfahrens fanden sie heraus, dass Anwärter auf einen Arbeitsplatz in Tests eine höhere Testmotivation und Anstrengung zeigten als bereits Angestellte. Zudem korrelierten die kognitiven Fähigkeiten signifikant mit den Faktoren des TAS. Arvey und Kollegen (1990) folgerten daraus, dass es individuelle Unterschieds- und Leistungsfaktoren gibt, die mit der Testmotivation in Verbindung stehen. Weitere Untersuchungen konnten die Ergebnisse von Arvey et al. (1990) bestätigen. Die Testbewertung kann tatsächlich einen Einfluss auf die Testleistung des Probanden haben, wenn diese von Bewertungsängsten oder einer ablehnenden Haltung beeinflusst werden (Salgado et al., 1996; Chan et al., 1997). So wurde ein positiver Zusammenhang zwischen Testleistung und Testmotivation gefunden, weshalb auch in solchen Tests die Testmotivation erhoben werden sollte.

An der Studie von Arvey et al. (1990) wurde jedoch in späteren Untersuchungen zu verschiedenen Aspekten Kritik geübt. So gründe das TAS nicht explizit auf einer spezifischen Motivationstheorie und sei unidimensional, weshalb es motivationale Konstrukte wie Instrumentalität und Erwartung nicht berücksichtigen kann (Sanchez et al., 2000). Bei nachfolgend entwickelten Verfahren zur Erfassung der Testmotivation wurden diese Kritikpunkte berücksichtigt, wie beispielsweise bei der Valence, Instrumentality, Expectancy Motivation Scale (VIEMS) von Sanchez et al. (2000) oder in einer Studie von Reeve und Lam (2007).

Weitere Untersuchungen zur Testmotivation greifen vor allem den Aspekt auf, welchen Einfluss die Testmotivation bei Tests ohne und mit Konsequenzen hat. Zunächst scheint es eindeutig zu sein, dass die Motivation für einen Test, der unmittelbare Konsequenzen für den zu Testenden hat, höher ist als ein Test ohne Konsequenzen. Wenn also die Ergebnisse in Tests ohne Konsequenzen mangels Motivation schlecht ausfallen (Wolf & Smith, 1995), kann das zu einer Verwechslung von Wissen und Motivation führen, wenn die Testmotivation nicht erfasst wird. Die Validität der Ergebnisse würde dadurch infrage gestellt. Jedoch sind die Ergebnisse bisheriger Studien, die diesen Gesichtspunkt berücksichtigt haben, mehrdeutig. So konnten beispielsweise Wolf und Smith (1995) zeigen, indem sie zwei Gruppen von Studenten/-innen zufällig auf zwei Tage verteilten und sie an einem Tag einen Test ohne und am anderen Tag einen Test mit Konsequenzen absolvieren ließen, dass die Konsequenzgruppe signifikant höhere Werte in der Testmotivation aufwies ($d = 1.45$). Zudem war in dieser Gruppe sowohl die Testleistung signifikant besser als auch die Korrelation zwischen Motivation und

Testleistung höher ($r = .35$ im Vergleich zu $r = .23$). Von einer Reduzierung der Motivation und der Testleistung bei Tests ohne Konsequenzen konnten unter anderem auch die Studien von Wise und DeMars (2005) und Wolf et al. (2005) berichten. Jedoch zeigten Sundre und Kitsantas (2004) in ihrer Untersuchung, dass ein Anstieg in der Testkonsequenz nicht unbedingt zu einer besseren Testleistung oder einer höheren Testmotivation führen muss. Sie fanden heraus, dass die Testmotivation kein signifikanter Prädiktor für die Testleistung in Examen mit Konsequenzen war, in Examen ohne Konsequenzen hingegen schon. Aufgrund der persönlichen Folgen des Examens mit Konsequenzen ist die Motivationsskala in dieser Bedingung hoch und weist eine niedrige Standardabweichung auf (Sundre & Kitsantas, 2004). Für die Gruppe ohne Konsequenzen spielt jedoch die Motivation, um die Testleistung vorherzusagen zu können, eine entscheidende Rolle. In dieser Bedingung fiel die Motivation niedriger aus, wies eine größere Variabilität auf, und die Testergebnisse waren niedriger. Dementsprechend kann eine niedrige Motivation die Testleistung in Tests ohne Konsequenzen untergraben, wohingegen eine hohe Motivation in Tests mit Konsequenzen die Testergebnisse nicht beeinflusst. Auch Baumert und Demmrich (2001), die eine kurze Form des OECD's Programme for International Student Assessment (PISA) verwendeten und dabei die Konsequenzen für den Test veränderten, konnten über die verschiedenen Bedingungen hinweg keinen Unterschied in der Testmotivation feststellen. Außerdem war sowohl die Anstrengung als auch die Testleistung in allen Bedingungen gleich. In der Untersuchung von Eklöf (2006), die einen Test, den schwedischen Trends in Mathematics and Science Study (TIMSS) 2003 verwendete, fand sich ebenfalls eine durchweg hohe Motivation der Testteilnehmer/-innen.

Darüber hinaus konnte in vielen Untersuchungen ein positiver Zusammenhang zwischen Testmotivation und Testleistung (unter anderem Arvey et al., 1990; Sundre & Kitsantas, 2004; Wise & DeMars, 2005; Wolf & Smith, 1995) bzw. ein nur teilweise signifikanter Zusammenhang zwischen Testergebnis und Testmotivation (Sanchez et al., 2000) gefunden werden. Bestätigt sich dieser Zusammenhang, so hat die Testmotivation auch einen Einfluss auf die Testvalidität (Sanchez et al., 2000).

1.3.3.2 Die Auswirkungen von Testmotivation in Leistungstests

Kann die Testmotivation einen Einfluss auf die Leistung in einem Leistungstest haben? Dieses Unterkapitel widmet sich einigen bisherigen Studien, die sich mit der Beantwor-

tung dieser Frage beschäftigt haben. Wie bereits im vorherigen Kapitel erwähnt, konnten mehrere Untersuchungen einen positiven Zusammenhang zwischen Testmotivation und Testleistung finden.

Normalerweise ist bei der Erhebung kognitiver Leistungstests davon auszugehen, dass sich alle Testpersonen in hinreichender Weise bemühen, ihr Leistungspotential auszuschöpfen. Sollten sie sich jedoch nicht ausreichend anstrengen, können die Testwerte die Fähigkeiten der Probanden nicht valide messen (Jacobs, 2007). Da in Leistungstests die Bearbeitungszeit meist feststeht, kann eine Steigerung der Testmotivation nur zu einer höheren Konzentration oder größeren Anstrengung führen. Es konnte jedoch gezeigt werden, dass sich durch Geldversprechen nicht nur die Testmotivation, sondern auch die kognitiven Testleistungen erhöhten. Aber ist der Rückschluss, dass sich durch ein Geldversprechen die Testmotivation erhöht und dadurch wiederum die Leistung wächst, richtig (Jacobs, 2009)? So konnten sowohl Gneezy und Rustchini (2000) in ihrer Untersuchung eine leistungssteigernde Wirkung von Geldanreizen für komplexere kognitive Aufgaben vorweisen als auch Kieffer und Goh (1981) die Intelligenztestleistungen von Unterschichtkindern durch Belohnungen für gute Leistungen deutlich steigern. Kieffer und Goh (1981) konnten zudem zeigen, dass eine Belohnung bei Mittelschichtkindern zu keiner weiteren IQ-Verbesserung führte.

Demnach ist nach Rindermann (2006) für das Testergebnis in Leistungstests die Intelligenz verantwortlich, vorausgesetzt, es liegt eine ausreichende Motivation vor, diesen Test zu bearbeiten (zitiert nach Jacobs, 2007). Diese Motivation ist meist dann vorhanden, wenn die Probanden von der Nützlichkeit einer Testung überzeugt werden konnten und zu einer ausreichenden Anstrengung bereit sind. In diesem Fall kann ein Geldversprechen zwar die Testmotivation noch zusätzlich erhöhen, diese wird jedoch nicht zu einer weiteren Leistungsverbesserung führen (Jacobs, 2007).

Die Testmotivation hat also dahingehend einen Einfluss auf die Leistung in einem Leistungstest, dass sie vorhanden sein muss, damit valide Ergebnisse, die die Fähigkeit einer Person erfassen soll, möglich sind. Ist sie andererseits bereits ausreichend vorhanden, steigert eine Erhöhung der Testmotivation das Leistungsergebnis nicht. Aus diesem Grund sollte in einem Leistungstest dafür Sorge getragen werden, dass die Testprobanden hinreichend motiviert sind, was meist durch die Standardinstruktionen, die den Sinn der Testung erklären, erreicht wird (Jacobs, 2007).

1.4 Testmotivation, Leistungsmotiv und Leistungstests

In diesem Kapitel wird zunächst das Modell von Covington und Kollegen (Covington & Omelich, 1991; Covington & Roberts, 1994) in die Modelle des RLMI-K/J und des AMG-S K-J eingebunden. Anschließend wird kurz auf einige empirische Befunde zum Zusammenhang zwischen Leistungsmotiv und Leistungen in Leistungstests eingegangen. Im Anschluss daran folgt der Versuch, den Zusammenhang zwischen Leistungsmotiv und Testmotivation in einem Leistungstest herauszustellen. In der einschlägigen Literatur finden sich einige Arbeiten über den Zusammenhang zwischen Leistungsmotiv und Leistungen in einem Leistungstest sowie über den Einfluss der Testmotivation auf die Leistungen. Die Beziehung zwischen Leistungsmotiv und Testmotivation wird jedoch in der gängigen Literatur kaum erwähnt. Aus diesem Grund werden hier, basierend auf den vorhandenen Erkenntnissen, Annahmen über den Einfluss des Leistungsmotivs auf die Testmotivation aufgestellt.

1.4.1 Zusammenführung der verschiedenen Leistungsmotive in das quadripolare Modell der Leistungsmotivation

In dieser Arbeit werden als Basis zur Erklärung des Leistungsmotivs die Annahmen des quadripolaren Modells der Leistungsmotivation von Covington et al. (Covington & Omelich, 1991; Covington & Roberts, 1994) zugrunde gelegt, weshalb zunächst die unterschiedlichen Leistungsmotivkomponenten der verwendeten Messverfahren, des RLMI-K/J und des AMG-S K-J, in das quadripolare Modell eingebunden werden.

Entsprechend der zugrundeliegenden Theorien gibt es sowohl im RLMI-K/J als auch im AMG-S K-J die Leistungsmotivkomponente HE. Das Leistungsmotiv HE kann mit dem Leistungsmotiv der Erfolgsorientierten (HL) von Covington et al. (Covington & Omelich, 1991; Covington & Roberts, 1994) gleichgesetzt werden, da auch dieses durch eine hohe Erfolgsorientierung bei gleichzeitig geringer Misserfolgsorientierung gekennzeichnet ist.

Gibt es im RLMI-K/J nur eine Komponente für die FM, so nimmt das AMG-S K-J eine Unterscheidung in zwei Komponenten, FMa und FMp, vor. Auch misserfolgsvermeidende (LH) Personen sind nach Covington et al. (Covington & Omelich, 1991; Covington & Roberts, 1994) misserfolgsängstlich, weisen dementsprechend neben einer starken Misserfolgsorientierung ebenfalls eine geringe Erfolgsorientierung auf, weshalb

sie Erfolgen wenig abgewinnen können. Dieses Motiv kann demzufolge mit den Motiven FM bzw. FMP gleichgesetzt werden.

Bei Personen mit einer starken FMA wirken deren Ängste hingegen nicht hemmend, sondern mobilisierend. So strengen sie sich sehr an, um einen Erfolg zu erreichen und einen Misserfolg zu vermeiden, genau wie es bei den übermotivierten (HH) Personen, gemäß dem Modell von Covington und Kollegen (Covington & Omelich, 1991; Covington & Roberts, 1994), der Fall ist. Auch diese Motive lassen sich folglich gleichsetzen.

Bisher sind die Motive FE und HM des RLMI-K/J unberücksichtigt geblieben. Personen mit einer starken FE sollten auch eine geringe Erfolgshoffnung aufweisen. Im Hinblick auf die Angst vor Misserfolg ist bei diesen Personen eine mittlere Ausprägung anzunehmen. Personen, bei denen hingegen die HM überwiegt, sollten über mögliche Misserfolge wenig besorgt sein, bezüglich ihrer Hoffnung auf Erfolg ist bei ihnen eine mittlere Ausprägung anzunehmen. Betrachtet man nun nicht die jeweils mittleren Ausprägungen der beiden Motive, sondern die verbleibenden und fasst diese zusammen, so erhält man das Motiv der Misserfolgsakzeptierenden (LL) nach Covington et al. (Covington & Omelich, 1991; Covington & Roberts, 1994). Personen mit diesem Motiv fühlen sich weder von Erfolgen angezogen noch sind sie in Leistungssituationen besorgt oder ängstlich. Dementsprechend lassen sich auch diese Motive in das quadripolare Modell einbinden.

1.4.2 Leistungsmotiv und Leistungstests

Es gibt eine Reihe von empirischen Befunden, die sich mit den Auswirkungen des Leistungsmotivs auf die Leistungen in Leistungstests beschäftigt haben. Hier soll nur eine kleine Auswahl vorgestellt werden. Grundsätzlich ist nach den bisher beschriebenen Theorien davon auszugehen, dass Personen mit einer überwiegenden HE bzw. HL in Leistungstests besser abschneiden als Personen mit überwiegender FM bzw. LH. So konnte beispielsweise Lowell (1952) zeigen, dass Versuchspersonen mit einem hohen Leistungsmotiv in einem verbalen Intelligenztest besser abschnitten als Personen mit geringerer Leistungsmotivation. Hinzu kam, dass in dieser Studie die hochmotivierten Testpersonen im Vergleich zu den Niedrigmotivierten einen schnelleren Lernfortschritt aufwiesen. Weitere Untersuchungen ergaben für die Bearbeitung von komplexen und schwierigen Aufgaben, welche ebenfalls in Intelligenztests vorkommen, ein ähnliches

Bild. Bei Schneider et al. (1973) etwa schnitten die Erfolgsorientierten im Vergleich zu den Misserfolgsmotivierten mit besseren Leistungsergebnissen ab. Aber auch die Befunde von French (1958) sowie French und Thomas (1958) deuten in diese Richtung. Sie konnten ebenfalls zeigen, dass die hochmotivierten Probanden gerade bei komplexen Problemlöseaufgaben länger nach einer Lösung suchten und deshalb bessere Leistungen erzielten. Sarason (1961; 1963) hat in verschiedenen Studien den Einfluss von FM in Testsituationen untersucht und kam dabei zu dem Schluss, dass die FM und das Testergebnis negativ zusammenhängen. Demnach könnte das schlechtere Abschneiden von Misserfolgsmotivierten in Tests daran liegen, dass die zur Lösung der Aufgaben notwendigen Kognitionen mit der Furcht in Tests interferieren (Sarason, 1963). Auch in der Studie von Stadler (2010), die die Beeinflussbarkeit der Intelligenztestleistung durch Testdurchführungssituation, Leistungsmotiv und implizite Intelligenztheorie untersuchte, erzielten die erfolgsmotivierten Versuchspersonen bessere Leistungen im Intelligenztest als die überwiegend Misserfolgsmotivierten. Zu dem gleichen tendenziellen Ergebnis kam auch Eiseler (2010) in ihrer Studie, in welcher sie den Einfluss von Misserfolgstress auf die Intelligenztestleistung in Abhängigkeit vom Leistungsmotiv untersuchte. Erfolgsorientierte erzielten hier tendenziell höhere Leistungen und zudem konnte ein signifikant negativer Zusammenhang zwischen der Intelligenztestleistung und der Furcht vor Misserfolg festgestellt werden.

In dieser Arbeit wird jedoch davon ausgegangen, dass der Zusammenhang zwischen Leistungsmotiv und Leistungen in Leistungstests über die Testmotivation vermittelt wird.

1.4.3 Der Zusammenhang zwischen Testmotivation, Leistungsmotiv und Leistungstests

Wie bereits erwähnt, wird zunächst davon ausgegangen, dass die Testmotivation durch die Standardinstruktionen ausreichend vorhanden ist. Doch welchen Einfluss hat nun das Leistungsmotiv auf die Testmotivation und wie wirkt sich dieses wiederum auf die Leistungen in einem Leistungstest aus? Diese Frage soll anhand von Annahmen und Schlussfolgerungen beantwortet werden. Zunächst wird erläutert, welchen Einfluss die verschiedenen Leistungsmotivtypen von Covington und Kollegen (Covington & Omelich, 1991; Covington & Roberts, 1994; siehe Kapitel 1.1.2.3) auf die Testmotivation in

Leistungstests haben. Im Anschluss daran werden Schlussfolgerungen über die Auswirkungen der Testmotivation auf die Leistungen in Leistungstests gezogen.

Bei den erfolgsorientierten (HL) Personen wird davon ausgegangen, dass sie sich anspruchsvollen Aufgaben, wie beispielsweise einem Intelligenz- bzw. Leistungstest, optimistisch sowie selbstbewusst stellen und zudem durch die Antizipation von Stolz bei erfolgreicher Bewältigung der Aufgabe motiviert werden (Brunstein & Heckhausen, 2006). Demnach ist anzunehmen, dass bei diesem Leistungsmotivtyp eine ausreichende Testmotivation für die Bearbeitung eines Leistungstests vorliegt und die Intelligenz für das Testergebnis verantwortlich ist (Jacobs, 2007).

Können misserfolgsvermeidende (LH) Personen Leistungssituationen, wie es bei einem Intelligenztest der Fall ist, nicht vermeiden, empfinden sie diese Aktivitäten als stark belastend (Rheinberg, 2006). Es ist aber davon auszugehen, dass Personen dieses Leistungsmotivtyps gerade bei Tests, die Aufgaben enthalten, die Rückschlüsse auf ihre Leistungen erlauben, hoch motiviert sein werden, um einen drohenden Misserfolg und das damit verbundene Schamgefühl zu vermeiden. Folglich ist auch bei diesem Leistungsmotivtyp anzunehmen, dass eine ausreichende Motivation für die Bearbeitung eines Leistungstests vorliegt. Sollten sie dennoch schlechtere Ergebnisse als die Erfolgsorientierten erzielen, so liegt dieses nicht an der mangelnden Testmotivation, sondern könnte, wie bereits Sarason (1963) vermutet, unter Umständen daran liegen, dass bei ihnen die zur Lösung der Aufgaben notwendigen Kognitionen mit der Furcht in Tests interferieren.

Bei den übermotivierten (HH) Personen sollte ebenfalls eine ausreichende Motivation für die Bearbeitung eines Leistungstests vorliegen. Erzielen sie dennoch schlechtere Ergebnisse als die Erfolgsorientierten, liegt es auch bei ihnen nicht an der mangelnden Testmotivation. Viel eher könnte ihre Einstellung, möglichst gut zu sein und alles möglichst detailliert zu bearbeiten, dazu führen, ineffektiv zu bleiben, gerade bei komplexen Aufgaben, wie sie in Intelligenztests gestellt werden. Auch der Umstand, sich in Testsituationen zu sehr unter Druck zu setzen, könnte dazu beitragen, dass sie ihre eigentlichen Fähigkeiten untergraben.

Die misserfolgsakzeptierenden (LL) Personen stehen Leistungssituationen, wie etwa einem Intelligenz- bzw. Leistungstests, eher gleichgültig gegenüber und bemühen sich nicht, ihre eigenen Leistungen zu erhöhen (Brunstein & Heckhausen, 2006). Bei diesem Leistungsmotivtyp ist davon auszugehen, dass keine ausreichende Motivation für die Bearbeitung eines Leistungstests vorliegt. Für das Testergebnis in einem solchem Test

könnte nicht die Intelligenz, sondern die mangelnde Testmotivation verantwortlich sein. Dementsprechend würde das Testergebnis nicht das widerspiegeln, was es eigentlich zu messen vorgibt.

Zusammenfassend lässt sich festhalten, dass die Testmotivation als situationsspezifische Komponente der Leistungsmotivation einen Einfluss auf die Leistungen in einem Leistungstest ausüben kann. Entsprechend der vorhergegangenen Überlegungen ist davon auszugehen, dass es misserfolgsakzeptierenden Personen (LL) an der Testmotivation in Leistungstests mangeln wird und sich dieses negativ auf ihre Leistungen in einem derartigen Test auswirken wird. Die anderen drei Leistungsmotivtypen sollten durch die Standardinstruktionen eines Intelligenz- bzw. Leistungstests und ihre persönlichen Einstellungen zu leistungsorientierten Situationen ausreichend motiviert sein. Bei ihnen sollte die Intelligenz für das Testergebnis in einem solchen Test verantwortlich sein.

2. Fragestellungen und Hypothesen

Mit der vorliegenden empirischen Arbeit soll der Einfluss untersucht werden, den die Testmotivation der Versuchspersonen auf die Leistungen in einem Leistungstest ausübt. Dabei wird zunächst überprüft, inwieweit sich die grundlegenden Leistungsmotive in ein gemeinsames Modell integrieren lassen, und ob die Testmotivation durch die verschiedenen Skalen erfasst werden kann. Anschließend werden die Leistungsmotivtypen mit der situationsspezifischen Testmotivation in Verbindung gebracht. Der Einfluss der Testmotivation auf die Leistungen des LPS-neu wird im Anschluss überprüft. Auch sollen die Zusammenhänge zwischen Intelligenz und Schulleistung, besuchter Schulart sowie Jahrgangsstufe untersucht werden.

2.1 Leistungsmotiv

In dieser Studie wird als Basis zur Erklärung des Leistungsmotivs das quadripolare Modell der Leistungsmotivation von Covington et al. (Covington & Omelich, 1991; Covington & Roberts, 1994) verwendet. Dabei wird angenommen, dass die Einbindung der Komponenten der in dieser Studie verwendeten Messinstrumente des Leistungsmotivs, das RLMI-K/J und das AMG-S K-J, in das Modell von Covington et al. (Covington & Omelich, 1991; Covington & Roberts, 1994) erfolgreich ist (siehe auch Kapitel 1.4.1). Wie die einzelnen Motivkomponenten dabei zusammenhängen, kann der Abbildung 2.1 entnommen werden.

Entsprechend der Grafik 2.1 wird angenommen, dass die beiden Motive des RLMI-K/J, HM und FE, miteinander korrelieren und sich zu einer Motivkomponente, LL, des quadripolaren Modells zusammenfügen lassen. Des Weiteren sollten die beiden Komponenten HE aus dem RLMI-K/J und dem AMG-S K-J miteinander korrelieren und zu HL zusammengefasst werden können. Darüber hinaus wird ebenfalls davon ausgegangen, dass die Motivkomponente FM des RLMI-K/J mit dem Motiv FMp des AMG-S K-J korreliert und sich diese beiden zu LH des quadripolaren Modells zusammenfassen lassen. Zudem sollte die Motivausprägung FMa des AMG-S K-J der Motivkomponente HH des quadripolaren Modells entsprechen.

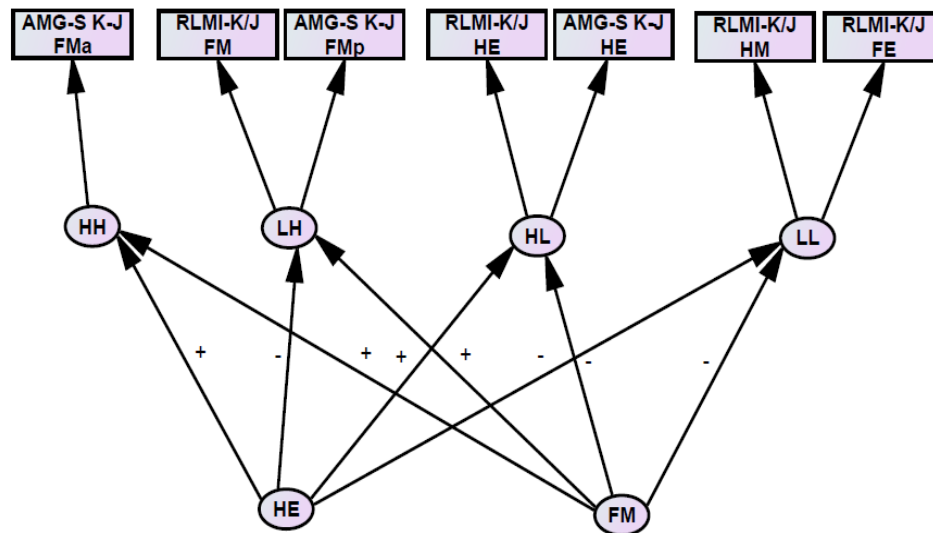


Abbildung 2.1

Zusammenhang zwischen RLMI-KJ bzw. AMG-S K-J und den Leistungsmotivkomponenten von Covington et al. (Covington & Omelich, 1991; Covington & Roberts, 1994)

Des Weiteren sind in der Abbildung 2.1 die beiden unabhängigen Konstrukte, HE und FM, sowie die Ausprägungen der Leistungsmotivtypen zu diesen unabhängigen Motiven eingezeichnet, wobei ein Plus-Zeichen einen hohen und ein Minus-Zeichen einen niedrigen Wert kennzeichnet.

Gemäß dieser Zusammenhänge und den oben genannten Annahmen lassen sich für die Leistungsmotivation folgende Hypothesen formulieren:

H1: Die Einbindung der Leistungsmotivkomponenten des RLMI-K/J und des AMG-S K-J in das quadrupolare Modell der Leistungsmotivation von Covington et al. (Covington & Omelich, 1991; Covington & Roberts, 1994) ist erfolgreich.

H2: Die Leistungsmotive HE und FM fungieren als unabhängige Konstrukte.

2.2 Testmotivation

In der zur Erfassung der Testmotivation zugrundegelegten Theorie, der Erwartungs-Wert-Theorie der Leistungsmotivation, gehen Eccles und Kollegen (1983) davon aus, dass die situationsspezifische Motivation sowohl von der Erwartung einer Person als auch vom Aufgabenwert beeinflusst wird (Eklöf, 2006).

Entsprechend der Modellannahmen (siehe dazu Kapitel 1.3.2) wird in dieser Studie angenommen, dass die Testmotivation in einem Leistungstest durch die beiden Subskalen – Erfolgserwartung und Aufgabenwert (siehe Abbildung 2.2) – beeinflusst wird. Gemäß dem Modell beeinflussen die beiden Skalen das Leistungsverhalten im Test direkt. Das Leistungsverhalten wiederum spiegelt sich in Ausdauer, Wahl und Leistung bzw. Anstrengung wider (Pohlmann et al, 2005). Die beiden oben genannten Skalen werden ihrerseits durch die Einstellungen und das allgemeine Leistungsverhalten beeinflusst. Da jedoch in dieser Arbeit der Einfluss auf die Testmotivation von Interesse ist, wird entsprechend davon ausgegangen, dass auch dieser Einfluss der beiden Skalen „Leistungsverhalten“ sowie „Einstellungen und Leistungen“ über die Testmotivation vermittelt wird.

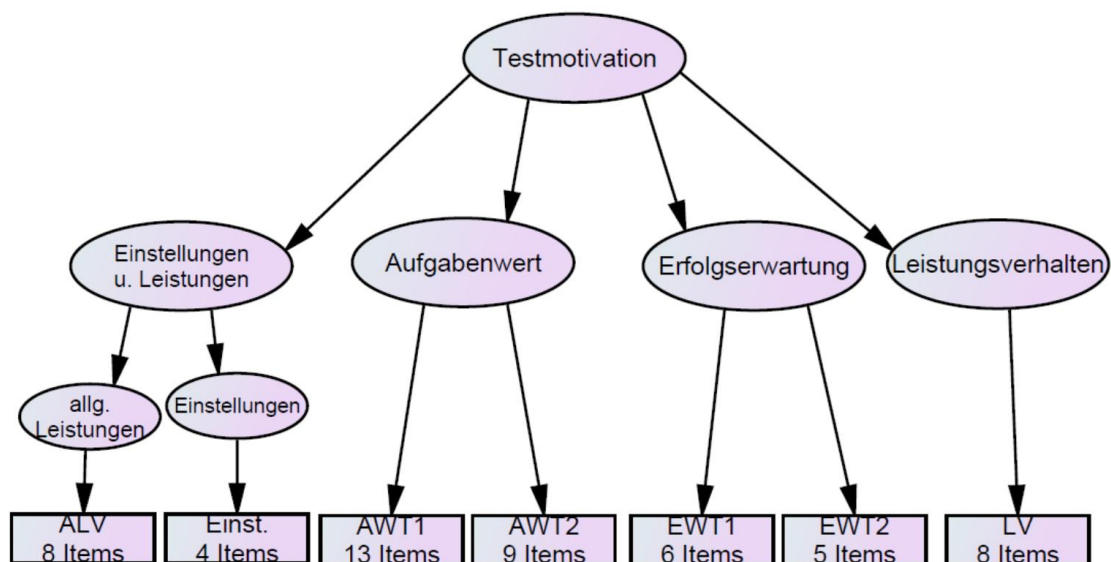


Abbildung 2.2

Subskalen der Testmotivation und die Anzahl der jeweils abgefragten Items

Es wird zudem davon ausgegangen, dass die Subskala „Erfolgserwartung“ sowohl vom Begabungsselbstkonzept als auch von der wahrgenommenen Aufgabenschwierigkeit direkt bestimmt wird (Pohlmann et al., 2005) und sich die Subskala „Aufgabenwert“ aus den vier Komponenten – Wichtigkeitskomponente, intrinsischer Wert, Nützlichkeitskomponente und Kosten – zusammensetzt (Eccles et al., 1983). Zudem sollte die Subskala „Einstellungen“, welche die individuellen Ziele, bisherige Erfolgs- und Misserfolgserfahrungen sowie deren Interpretationen und auch das kulturelle Milieu, erfasst, für die vollständige Erhebung der Testmotivation von Bedeutung sein. Diese Subskala

bildet zusammen mit den „allgemeinen Leistungen“ die Skala „Einstellungen und Leistungen“.

Aus den Modellannahmen von Eccles und Kollegen (1983) sowie weiteren Überlegungen lässt sich folgende Hypothese zu den Subskalen der Testmotivation aufstellen:

H3: Die Testmotivation kann mit den verschiedenen Subskalen der Theorie von Eccles et al. (1983) erfasst werden.

2.3 Intelligenz

Das Drei-Ebenen-Modell von Carroll (1993) bietet den theoretischen Rahmen für das in dieser Studie zur Messung der Intelligenz verwendete LPS-neu von Kreuzpointner (2010), welcher ebenfalls eine Unterscheidung in Stratum I, II und III vornimmt. In Analysen zur Struktur des LPS-neu, welche im Kontext des Drei-Ebenen-Modells von Carroll (1993) stattfanden, konnte Kreuzpointner (2010) zeigen, dass die Konstruktvalidität gegeben ist. Zudem deuten die Ergebnisse der psychometrischen Eigenschaften von Kreuzpointners Untersuchung aus dem Jahr 2010 darauf hin, dass das LPS-neu auch im Hinblick auf die Gütekriterien sowie die Verteilungen ein brauchbares Verfahren ist.

Kreuzpointners Ergebnissen zufolge können durch das LPS-neu mithilfe der „elf unterschiedlichen Subtests vier Faktoren kognitiver Kompetenz zweiter Ordnung (kristalline und fluide Intelligenz, visuelle Wahrnehmung und kognitive Geschwindigkeit) und ein Faktor dritter Ordnung (Allgemeine Intelligenz, g) gemessen werden“ (2010, S. 162). Die verschiedenen Ordnungen können dem Strukturmodell (siehe Kapitel 3.4.6.1) ebenfalls entnommen werden.

Für die Struktur, welche dem LPS-neu zugrundeliegt, lässt sich folgende Hypothese bilden:

H4: Die formulierte Struktur des LPS-neu von Kreuzpointner (2010) lässt sich replizieren.

2.4 Zusammenhänge

Zunächst wird der Zusammenhang zwischen Leistungsmotiv und Testmotivation dargestellt. Anschließend soll, ausgehend davon, dass der Zusammenhang zwischen Leis-

tungsmotiv und Leistungstests über die Testmotivation vermittelt wird, auf die Beziehung von Testmotivation zu den Leistungstestergebnissen des LPS-neu eingegangen werden. Zum Schluss wird der Zusammenhang zwischen Intelligenz und Schulleistung sowie zwischen Intelligenz und besuchter Schulart bzw. Jahrgangsstufe erläutert.

2.4.1 Leistungsmotiv und Testmotivation

Wie bereits in Kapitel 1.4.1 ausführlich dargestellt, wird zunächst postuliert, dass die Testmotivation durch das Leistungsmotiv, aus welchem sich die Testmotivation als situationsspezifische Komponente ableiten lässt, beeinflusst wird. Demnach lassen sich aus den vier, von Covington und Kollegen (Covington & Omelich, 1991; Covington & Roberts, 1994) vorgeschlagenen Leistungsmotivtypen überprüfbare Hypothesen über deren jeweiligen Einfluss auf die Testmotivation ableiten. Dieser Zusammenhang ist in der nachfolgenden Hypothese formuliert:

H5: Versuchspersonen, bei denen das Leistungsmotiv HH, HL bzw. LH hoch ausgeprägt ist, weisen eine ausreichende Testmotivation für die Bearbeitung eines Leistungstests auf.

Zu klären bleibt jedoch, sofern sich die oben genannte Hypothese bestätigt, ob es Unterschiede, abhängig vom Leistungsmotivtyp, in der Höhe der vorliegenden Testmotivation gibt. Darüber hinaus wäre es interessant zu analysieren, ob bei Kenntnis des Leistungsmotivs auf die Testmotivation geschlossen werden kann. Aufgrund dessen sollen in dieser Untersuchung die folgenden Fragestellungen überprüft werden:

F1: Gibt es Unterschiede, abhängig vom Leistungsmotivtyp, in der Höhe der vorliegenden Testmotivation (HH bzw. HL höher als LH bzw. LL)?

F2: Kann bei Kenntnis des Leistungsmotivs auf die Testmotivation geschlossen werden?

2.4.2 Testmotivation und Leistungen in einem Leistungstest

Gemäß den in Kapitel 1.3.3.2 dargestellten empirischen Studien wird zunächst die folgende Hypothese über den Zusammenhang zwischen Testmotivation und Leistungen im LPS-neu postuliert:

H6: Die Testmotivation hat einen positiven Einfluss auf die Leistungen des LPS-neu.

Entsprechend der dort vorgestellten Studien hat die Testmotivation dahingehend einen Einfluss auf die Leistung in einem Leistungstest, dass sie vorhanden sein muss, damit valide Ergebnisse, die die Fähigkeit einer Person erfassen sollen, möglich sind. Ist sie bereits ausreichend vorhanden, steigert eine Erhöhung derselben das Leistungsergebnis nicht. In diesem Fall ist nicht die Motivation, sondern die Intelligenz für das Testergebnis in Leistungstests verantwortlich (Jacobs, 2007). Dementsprechend sollte eine geringe Testmotivation auch zu geringeren Intelligenztestwerten führen. Steigt die Testmotivation in einem Test an, müsste dieses auch zu besseren Ergebnissen der Leistungen in diesem Test führen, bis auch eine Steigerung der Testmotivation aufgrund kognitiver Grenzen keine Verbesserungen der Intelligenzergebnisse mehr nach sich zieht. Grafisch lässt sich der Zusammenhang zwischen Intelligenztestergebnissen und Testmotivation idealtypisch wie folgt darstellen:

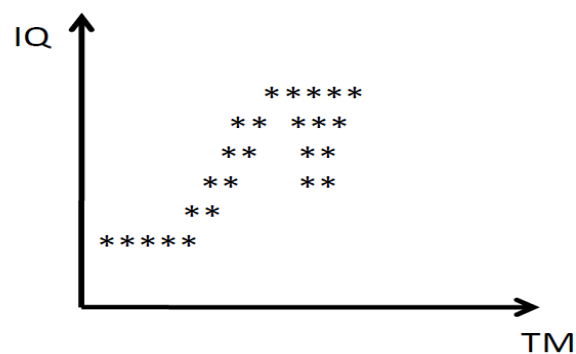


Abbildung 2.4

Zusammenhang zwischen Intelligenztestergebnissen und Testmotivation

Ausgehend von der Annahme, dass die Testmotivation und die Leistungen in einem Leistungstest zusammenhängen, werden nachstehend folgende Hypothesen aufgestellt:

H7: Der Zusammenhang zwischen Testmotivation und Intelligenz ist durch einen Leistungsanstieg aufgrund höherer Testmotivation begründet. Bei hoher Testmotivation tendiert die Korrelation gegen Null.

H8: Sind die Versuchspersonen für die Bearbeitung des LPS-neu ausreichend motiviert, dann ist die Intelligenz für das Testergebnis in diesem Test verantwortlich.

2.4.3 Schulleistung, besuchte Schulart bzw. Jahrgangsstufe und Intelligenz

Wie bereits detailliert in Kapitel 1.2.4 erläutert, kann beispielsweise Rost (2009) zufolge die allgemeine Intelligenz als bester Einzelprädiktor für die Schulleistung angesehen werden. Darüber hinaus lassen sich nach Kühn (1987), bei Kenntnis der allgemeinen Intelligenz, Zensuren in den Hauptfächern, etwa Mathematiknoten, besser vorhersagen als Nebenfächernoten. Auch Funke und Vaterrodt-Plünnecke (2004) weisen darauf hin, dass die Korrelation sowohl bei Mathematik- als auch bei Deutschnoten mit Intelligenztestwerten am höchsten ist.

Entsprechend der vorliegenden empirischen Studien zur Schulleistung und Intelligenz lassen sich für diese Zusammenhänge nachfolgende Hypothesen formulieren:

H9: Schüler/-innen mit einer guten Schulleistung weisen höhere Werte in der allgemeinen Intelligenz des LPS-neu auf als Schüler/-innen mit einer schlechteren Schulleistung.

H10: Die Hauptfächernoten der Versuchspersonen eignen sich als Prädiktor für die allgemeine Intelligenz besser als die Noten aus den Nebenfächern.

Zudem wäre es darüber hinaus interessant zu prüfen, inwieweit sich auch in dieser Studie die Zusammenhänge zwischen besuchter Schulart und Intelligenz, entsprechend der Untersuchung von Wechsler aus dem Jahre 1964, replizieren lassen, in welcher die schulische Ausbildung, gemessen in Jahren, mit den Intelligenztestergebnissen für eine Erwachsenenstichprobe hoch korrelierte. Des Weiteren soll der Frage nachgegangen werden, ob die besuchte Jahrgangsstufe einen Einfluss auf die Intelligenztestergebnisse ausübt. Ebenso wird untersucht, ob auch in dieser Studie ein Rückgang der Korrelation zwischen Schulleistung und Intelligenz mit steigendem Alter und Ausbildungsniveau durch das Vorwissen einer Person, wie Helmke und Weinert (1997) vermutet haben,

erklärt werden kann. Demzufolge sollte der Schulerfolg in den höheren Klassen bzw. am Ende einer Schullaufbahn vorrangig vom Wissen, welches über die Schuljahre hinweg aufgebaut wurde, abhängig sein. Aus diesem Grund werden die anschließenden Hypothesen postuliert:

H11: Die besuchte Schulart korreliert mit den Intelligenztestergebnissen.

H12: Die besuchte Jahrgangsstufe hat einen Einfluss auf die Intelligenztestergebnisse.

H13: Die Korrelation zwischen Schulleistung und Intelligenz sinkt mit steigendem Alter und Ausbildungsniveau.

3. Methode

Um die in Kapitel 2 dargestellten Fragestellungen und Hypothesen zu überprüfen, wurden verschiedene sogenannte „Papier- und Bleistift“-Fragebögen verwendet. Dabei wurden die Testleistungen mithilfe des überarbeiteten Leistungsprüfsystem (LPS-neu) erfasst. Weiterhin fanden folgende Tests und Fragebögen in der Untersuchung Anwendung:

- Fragebogen zur Erfassung der demographischen Daten
- Regensburger Leistungs-Motiv-Inventar für Kinder und Jugendliche (RLMI-K/J)
- Leistungsmotivationsinventar für Kinder und Jugendliche - kurz (AMG-S K-J)
- Fragebögen zur Erfassung der Testmotivation

In den anschließenden Abschnitten wird die Stichprobe beschrieben, der Verlauf des Experiments vorgestellt sowie die statistischen Verfahren und die Untersuchungsmethode erklärt. Darüber hinaus wurden die verschiedenen Komponenten in ein Strukturmodell integriert, das ebenfalls vorgestellt werden soll.

3.1 Beschreibung der Stichprobe

An dieser Untersuchung nahmen insgesamt 361 Schüler und Schülerinnen aus neunten und zehnten Klassen teil. Dabei verteilte sich die Stichprobe auf vier staatliche bayerische Schulen, ein Gymnasium, zwei Realschulen sowie eine Grund- und Mittelschule. Am Ludwigsgymnasium in Straubing wurde die Studie an jeweils drei neunten und zehnten Klassen durchgeführt. An der Agnes-Bernauer-Schule Realschule für Mädchen in Augsburg wurden wie am Gymnasium auch drei neunte und drei zehnte Klassen untersucht. Die Knabenrealschule Heilig Kreuz in Donauwörth stellte fünf Klassen der neunten Jahrgangsstufe zur Verfügung und an der Grund- und Mittelschule Schwarzach nahmen zwei neunte und eine zehnte Klasse teil. Somit verteilte sich die gesamte Stichprobe schließlich auf insgesamt 20 Klassen. Die Versuchspersonen nahmen freiwillig an der Untersuchung teil und erhielten bei Abschluss der Datenauswertung mithilfe ihres persönlichen Versuchspersonencodes ein Feedback über ihre Intelligenzleistungen im LPS-neu.

3.1.1 Besuchte Schulart und Klassen

Von den 361 Versuchspersonen besuchten zum Testzeitpunkt 135 das Ludwigsgymnasium in Straubing. Diese Probanden stellen mit 37.4 % den größten Personenkreis der teilnehmenden Schüler/-innen dar. Die Agnes-Bernauer-Schule Realschule für Mädchen in Augsburg nahm mit 98 Schülerinnen und einem Stichprobenanteil von 27.1 % teil. 90 Probanden, also 24.9 %, besuchten die Knabenrealschule Heilig Kreuz in Donauwörth. Die Grund- und Mittelschule Schwarzach stellt mit 10.5 % und 38 Probanden den kleinsten Stichprobenteil dar.

Zudem wurde die besuchte Schulart zum Zeitpunkt der Erhebung und im vorangegangenen Schuljahr erhoben (siehe Tabelle 3.1). An der Grund- und Mittelschule wurden diese Daten nicht erfasst, da hier von einem Wechsel auf eine andere Schule nicht ausgegangen wurde.

Tabelle 3.1

Besuchte Schulart im Schuljahr 2009/2010

Gymnasium	Häufigkeit	Prozent
keine Angabe	3	2.2
Gymnasium	126	93.3
Realschule	6	4.4
Gesamt	135	100.0
Knabenrealschule	Häufigkeit	Prozent
Gymnasium	1	1.1
Realschule	89	98.9
Gesamt	90	100.0
Mädchenrealschule	Häufigkeit	Prozent
Gymnasium	2	2.0
Realschule	96	98.0
Gesamt	98	100.0

Nur sechs Schüler/-innen besuchten ein Schuljahr vorher nicht das Gymnasium, sondern die Realschule und umgekehrt waren drei Schüler/-innen nicht an der Realschule, sondern am Gymnasium, wobei drei Schüler/-innen zur besuchten Schulart im vergangenen Schuljahr keine Angabe gemacht haben. Mit insgesamt 9 Schüler/-innen beträgt die Quote derjenigen, die die Schule gewechselt haben, lediglich 2.49 % der Gesamtstichprobe. Von den insgesamt 361 Probanden befanden sich zum Zeitpunkt der Untersuchung 250 in der neunten und 111 in der zehnten Jahrgangsstufe. Somit ist der Anteil

der Schüler/-innen, die zum Zeitpunkt der Untersuchung die neunte Klasse besucht haben, mit 69.3 % höher als der der zehnten Klassen, der lediglich 30.7 % beträgt.

3.1.2 *Geschlecht nach Schule und Klasse*

Insgesamt waren von den 361 Probanden 207 (57.3 %) männlich und 154 weiblich (42.7 %). Die Geschlechter verteilen sich wie folgt auf die teilnehmenden Schulen:

Tabelle 3.2

Verteilung der Geschlechter

Schule	N	männlich (in %)	weiblich (in %)
Gymnasium	135	99 (73.3 %)	36 (26.7 %)
Mittelschule	38	18 (47.4 %)	20 (52.6 %)
Mädchenrealschule	98	-	98 (100 %)
Knabenrealschule	90	90 (100 %)	-
Gesamt	361	207 (57.3 %)	154 (42.7 %)

Das deutliche Übergewicht der Schüler an der Gesamtstichprobe ergibt sich somit aus der hohen Teilnehmerzahl der männlichen Probanden des Ludwigsgymnasiums. Von den 250 Testpersonen der neunten Klassen waren 97 weiblich und 153 männlich. In dieser Klassenstufe wurden dementsprechend lediglich 38.8 % weibliche Probanden und 61.2 % männliche Probanden untersucht. Von den 111 Probanden der zehnten Jahrgangsstufe waren 53 (47.7 %) männlich und 58 (52.3 %) weiblich. In den neunten Klassen wurden mehr männliche und in den zehnten Klassen mehr weibliche Probanden getestet.

3.1.3 *Alter*

Das durchschnittliche Alter der getesteten Schüler/-innen liegt bei 15.34 Jahren und reicht von 14 bis 19 Jahren. Die Standardabweichung des Alters in dieser Stichprobe liegt bei 0.98. Mit einem Anteil von 42.4 % bildet die Altersgruppe der 15-Jährigen mit insgesamt 153 Versuchspersonen den größten Anteil (siehe Tabelle 3.3).

Tabelle 3.3

Verteilung des Alters nach ganzen Jahren

	Häufigkeit	Prozent
14	67	18.6
15	153	42.4
16	107	29.6
17	20	5.5
18	13	3.6
19	1	0.3
Gesamt	361	100.0

Werden zusätzlich auch die Monate betrachtet, so liegt das durchschnittliche Alter bei 15.81 Jahren, also etwas höher als bei der reinen Betrachtung des Alters nach Jahren. Aufgrund fehlender Angaben konnte das genaue Alter nach Monaten nur bei 350 Versuchspersonen berechnet werden. Die Spannweite beträgt auch hier fünf Jahre. Für die Verteilung des Alters, genau nach Monaten, liegt die Standardabweichung bei 0.92. Wird zusätzlich zum Alter in ganzen Jahren auch die Jahrgangsstufe betrachtet, sind die Schüler/-innen der neunten Klassen ($M = 15.02$; $SD = .78$) im Mittel ein Jahr jünger als die Schüler/-innen der zehnten Klassen ($M = 16.07$; $SD = 1.00$). Getrennt nach Geschlecht und Alter in ganzen Jahren liegt der Altersmittelwert der männlichen Probanden bei 15.18 Jahren, mit einer Standardabweichung von 0.83 und einer Spannweite von fünf Jahren. Die 154 weiblichen Probanden weisen einen Altersmittelwert von 15.56 Jahren auf, mit einer Standardabweichung von 1.11 und einer Spannweite von vier Jahren. Werden auch die Monate berücksichtigt, so liegt der Altersmittelwert der 199 männlichen Probanden bei 15.65 Jahren, mit einer Standardabweichung von 0.80. Die Spannweite liegt hier ebenfalls bei fünf Jahren. Bei den 151 Schülerinnen liegt der Altersmittelwert bei 16.02 Jahren, mit einer Standardabweichung von 1.03 und einer Spannweite von 4.58 Jahren. Somit wird deutlich, dass die weiblichen Probanden in der vorliegenden Stichprobe etwas älter sind als die männlichen.

3.1.4 Schulfächer und Schulnoten

Es wurden an allen Schulen die Schulnoten zu Haupt- und Nebenfächern erhoben. Die folgende Auflistung, getrennt nach Schularten, gibt an, zu welchen Schulfächern jeweils die Schulnoten erfasst wurden:

Gymnasium

Hauptfächer:	Deutsch, Mathematik, Englisch, 2. Fremdsprache
Nebenfächer:	Physik, Chemie, Erdkunde, Geschichte, Musik, Biologie, Religionslehre bzw. Ethik, Wirtschaft und Recht, Sport

Realschulen

Hauptfächer:	Deutsch, Mathematik, Englisch, 4. Prüfungsfach bzw. Wahlpflichtfach
Nebenfächer:	Erdkunde, Geschichte, Musik, Physik bzw. Chemie, Religionslehre bzw. Ethik, Sozialkunde, Sport

Grund- und Mittelschule

Hauptfächer:	Deutsch, Mathematik, Englisch
Nebenfächer:	GSE (Geschichte, Sozialkunde, Erdkunde), PCB (Physik, Chemie, Biologie), Musik, Religionslehre bzw. Ethik, Sport

Aufgrund zu vieler fehlender Daten wurden einige Fächer nicht berücksichtigt: am Gymnasium die Fächer Wirtschaft und Recht sowie Erdkunde, an der Knabenrealschule Musik und Sozialkunde, an der Mädchenrealschule das Wahlpflichtfach sowie Sozialkunde und an der Grund- und Mittelschule das Fach Musik. Waren in den Fächern Physik bzw. Chemie zwei Noten angegeben, wurde daraus die Durchschnittsnote berechnet und diese in den weiteren Berechnungen berücksichtigt.

3.1.5 Klasse wiederholt

Zusätzlich wurde erfasst, wie viele Schüler/-innen die zum Testzeitpunkt besuchte Klasse bereits wiederholten. Diese Daten wurden an der Grund- und Mittelschule nicht erfragt, da hier davon ausgegangen werden kann, dass nur ein sehr geringer Teil eine Klasse wiederholen muss. Die folgende Tabelle stellt dar, wie viele Schüler/-innen sich im Schuljahr 2010/2011 sowie im Schuljahr zuvor in den jeweiligen Klassen befanden und wie viele, getrennt nach Schulart, die Klassen wiederholen mussten.

Tabelle 3.4

Übersicht über die Verteilung der Klassenwiederholer

Schule		Klasse 09/10	Klasse 10/11	Wiederholer
Gymnasium	8. Klasse	66		
	9. Klasse	3	69	3
Knabenrealschule	8. Klasse	80		
	9. Klasse	10	90	10
Mädchenrealschule	8. Klasse	59		
	9. Klasse	6	65	
	keine Angabe	1		6
Gymnasium	9. Klasse	59		
	10. Klasse	7	66	7
Mädchenrealschule	9. Klasse	25		
	10. Klasse	5	32	
	keine Angabe	2		5
Gesamt		323	323	31

Entsprechend der Tabelle 3.4 mussten insgesamt 31 Probanden entweder die neunte oder die zehnte Klasse im Schuljahr 2010/2011 wiederholen. Von diesen 31 Wiederholern besuchten zum Testzeitpunkt 10 Schüler/-innen das Gymnasium, 10 Schüler die Knabenrealschule und 11 Schülerinnen die Mädchenrealschule. Drei Versuchspersonen hatten keine Angabe dazu gemacht, welche Klasse sie im Schuljahr 2009/2010 besuchten. In Kapitel 3.5 findet sich die Überprüfung, inwiefern sich die Wiederholer in ihren Leistungen im LPS-neu von den Nicht-Wiederholern unterscheiden.

3.2 Untersuchungsablauf

Bevor mit der Untersuchung, die an bayerischen Schulen durchgeführt werden sollte, begonnen werden konnte, wurde eine Genehmigung des Bayerischen Staatsministeriums für Unterricht und Kultus benötigt. Nach Erhalt der Genehmigung wurden insgesamt 16 Schulen entweder per E-Mail oder telefonisch kontaktiert und über die Untersuchung informiert. Lediglich vier Schulen waren mit der Testung einverstanden und wurden über die genaue Vorgehensweise entweder persönlich oder per E-Mail informiert. Im Anschluss daran wurden Termine für die eigentliche Testung vereinbart, welche dann im Januar und Februar 2011 stattfanden.

Da die meisten der Versuchspersonen unter 18 Jahre alt waren, wurde für die Teilnahme an der Untersuchung zusätzlich eine Einverständniserklärung der Eltern benö-

tigt, die am Testtag vorliegen musste. Zu diesem Zweck wurde ein Elternanschreiben entworfen, in dem sowohl die Eltern als auch die Schüler/-innen über den Inhalt und den Ablauf der Testung informiert wurden. Zudem wurde darum gebeten, das beigelegte Blatt zu den soziodemographischen Daten auszufüllen und am Testtag beim Versuchsleiter abzugeben. Darüber hinaus wurden an der Grund- und Mittelschule aus Zeitgründen auch die beiden Fragebögen zur Erfassung der Leistungsmotivation bereits vor der Testung ausgefüllt und am Testtag mitgebracht.

Am eigentlichen Testtag wurde der Versuchsleiter entweder von Lehrer/-innen bzw. vom Schulleiter in die Klassen gebracht. An der Mädchenrealschule konnten keine Lehrer/-innen bereitgestellt werden, da die Testungen während Vertretungsstunden stattfanden, sodass der Versuchsleiter alleine in die Klassen ging. Nachdem die Lehrer/-innen die Schüler/-innen um Ruhe gebeten hatten, wurden zunächst die Einverständniserklärungen eingesammelt und überprüft, sodass nur die Schüler/-innen an der Untersuchung mitmachen konnten, die auch eine unterschriebene Einverständniserklärung der Eltern vorwiesen. Da die Teilnahme freiwillig war, gab es in den Klassen auch einige Schüler/-innen, die nicht teilnehmen wollten bzw. durften. Diese wurden entweder für andere Tätigkeiten aus dem Klassenraum geführt oder zur Stillarbeit im Klassenzimmer aufgefordert.

Nachdem die Testteilnahme aller Schüler/-innen einer Klasse geklärt war, wurde die Untersuchung vom Versuchsleiter kurz vorgestellt. Zudem wurde nochmals darauf verwiesen, dass es sich um einen anonymen Test handle, bei dem keine Rückschlüsse auf die einzelnen Versuchspersonen vorgenommen werden, sodass die Teilnehmer/-innen die Fragebögen entsprechend ihrer Meinung ehrlich beantworten könnten.

Bevor mit dem eigentlichen Test, dem LPS-neu, begonnen wurde, verteilten der Versuchsleiter bzw. die jeweils anwesenden Lehrer/-innen die Fragebögen zur Erfassung der Testmotivation „vor dem Test“. Sobald alle Schüler/-innen mit dem Ausfüllen der Bögen fertig waren, wurden diese wieder eingesammelt und daraufhin die LPS-neu Testbögen, Version A bzw. B, an alle Schüler/-innen ausgehändigt. Bei der Verteilung wurde darauf geachtet, dass direkt nebeneinander sitzende Schüler/-innen unterschiedliche Versionen des Tests erhielten, um ein mögliches Abschreiben zu verhindern. Nachdem die Eingangsinstruktion, die das LPS-neu noch einmal kurz erklärte, vom Versuchsleiter laut vorgelesen wurde, konnte mit dem eigentlichen Test begonnen werden. Nach Beendigung des LPS-neu wurden alle Testbögen eingesammelt.

Dann wurden die Schüler/-innen gebeten, zunächst den bereits ausgeteilten Fragebogen zur Testmotivation „nach dem Test“ auszufüllen und anschließend mit der Beant-

wortung der Fragebögen zur Leistungsmotivation zu beginnen, wobei hier die Reihenfolge der Bearbeitung freigestellt wurde. Im Gegensatz zu den anderen Schulen wurde am Gymnasium das Blatt zu den soziodemographischen Daten erst zu diesem Zeitpunkt ausgefüllt. Nachdem alle Schüler/-innen fertig waren, wurden die Fragebögen eingesammelt.

Nach Beendigung der gesamten Testung wurden Fragen der Schüler/-innen zur Untersuchung beantwortet sowie nochmals auf die Ergebnismeldung nach Ende der Dateneingabe verwiesen. Am Abschluss bedankte sich der Versuchsleiter für die Teilnahme.

Nachdem alle Daten der Untersuchung eingegeben und ausgewertet waren, erhielten die Schulleiter und die einzelnen Klassen ein Schreiben mit dem Hinweis, dass die Testergebnisse des Leistungstests im Internet abgefragt werden können.

3.3 Statistische Datenauswertung – verwendete Verfahren

Die Auswertung der Daten wurde mittels der Version 18.0 des „Statistical Package for the Social Sciences“ (SPSS) und der „Analysis of Moment Structures“ (AMOS) sowie „Microsoft Office Excel 2007“ vorgenommen.

Vor Beginn der Auswertungen wurden die Variablen hinsichtlich ihres Skalenniveaus geprüft, um für die in Kapitel 2 genannten Zusammenhänge geeignete Korrelationsmaße zu finden.

Um die Stärke der Zusammenhänge, beispielsweise einzelner Skalen der Testmotivation miteinander bzw. der Testmotivation mit den Leistungen im LPS-neu berechnen zu können, wurden Korrelationen nach Bravais-Pearson verwendet. Darüber hinaus wurde der Zusammenhang zwischen den Haupt- bzw. Nebenfächernoten mit den Leistungen im LPS-neu mithilfe einer linearen Regressionsanalyse berechnet. Ziel dieser Analyse ist es, den Einfluss einer erklärenden Variable auf die abhängige Variable zu untersuchen. Die einzelnen Haupt- bzw. Nebenfächernoten stellten dabei die unabhängige und der Gesamtscore des LPS-neu die abhängige Variable dar. Als Maß für die Güte der Anpassung durch die Regressionsgerade gilt das Bestimmtheitsmaß R -Quadrat (R^2). Dabei entspricht R^2 dem Quotienten aus dem erklärten Teil der Varianz und der Gesamtvarianz. Die Regressionsanalyse kann nur durchgeführt werden, wenn die abhängige Variable metrisch skaliert ist.

Zudem wurde bei den Berechnungen der Leistungsmotive eine multidimensionale Skalierung (MDS-Alscal) vorgenommen. Bei der MDS werden Positionierungsanalysen durchgeführt, wobei zwischen den Objekten wahrgenommene globale Ähnlichkeiten, erfragt werden (Backhaus et al., 2008). Durch diese berechneten Ähnlichkeiten können mithilfe der MDS die zugrundeliegenden Wahrnehmungsdimensionen abgeleitet und anschließend die Objekte mittels dieser Dimensionen im Raum positioniert und grafisch dargestellt werden (Backhaus et al., 2008).

Für Vergleiche der Merkmale bzw. zur Überprüfung von Mittelwertsunterschieden zwischen zwei Merkmalen, zum Beispiel zwischen Version A und B des LPS-neu, wurden *t*-Tests für unabhängige Stichproben verwendet. Beim Vergleich der Ausprägung der Testmotivation mit den Leistungsmotivclustern wurden *t*-Tests bei einer Stichprobe berechnet. *T*-Tests wurden dann verwendet, wenn die Voraussetzungen für ihre Anwendung – das Vorliegen einer Normalverteilung und gleiche Varianzen (siehe Kapitel 3.4) – für die verschiedenen Messinstrumente als gegeben angenommen werden konnte. Dabei diente der Levene-Test, bei *t*-Tests für unabhängige Stichproben, zur Überprüfung der Varianzhomogenität. Sofern die Varianzen nicht gleich waren, fand der Welch-Test für die Berechnung von Mittelwertsunterschieden Anwendung. Die dabei angegebenen Effektstärken wurden bei unabhängiger Stichprobe und mit unterschiedlichen Stichprobengrößen mithilfe von Hedges *g* – die mittlere Differenz beider Mittelwerte wird durch die gepoolte Standardabweichung geteilt – berechnet. Bortz (1999) und Cohen (1988) geben Vorschläge für die Interpretation der Höhe der Effekte. So bezeichnen sie Werte um .20 eher als schwach, Werte um .50 als mittleren und Werte mit .80 als starken Effekt.

Sofern mehrere Mittelwertsunterschiede gleichzeitig überprüft werden sollten, beispielsweise beim Einfluss der Leistungsmotivcluster auf die Testmotivation, wurden diese mithilfe multivariater Varianzanalysen berechnet. Auch hier wurde, um den Einfluss der einzelnen Variablen angeben zu können, eine Effektstärke berechnet. Dabei kann zwischen Eta-Quadrat (η^2) und dem partiellen Eta-Quadrat ($_{\text{partial}}\eta^2$) unterschieden werden (Cohen, 1973; zitiert nach Kreuzpointner, 2010). Für diese Studie wurde das partielle Eta-Quadrat verwendet, welches die Höhe der Variation der abhängigen Variable auf die Unterschiede der unabhängigen Variablen beschreibt. Auch für dieses Effektmaß gibt Cohen (1988; zitiert nach Kreuzpointner, 2010) Interpretationshilfen. So bezeichnet er ein partielles Eta-Quadrat von $>.01$ als einen kleinen Effekt, einen Wert von $>.06$ als mittleren und alles über .14 als großen Effekt. Waren die Mittelwertsunter-

schiede signifikant, wurde der Scheffé Post-Hoc-Test angewandt, der den Vergleich von Mittelwertskombinationen ermöglicht.

Für die Überprüfung der einzelnen Strukturgleichungsmodelle und zur Prüfung der Modellgültigkeit wurden verschiedene Indizes herangezogen, die mithilfe von AMOS berechnet wurden. Dabei diente als Schätzmodell das robusteste Verfahren, die Maximum-Likelihood-Methode (Blunch, 2008; zitiert nach Kreuzpointner, 2010). Um dieses Schätzmodell anwenden zu können, sollte die Stichprobe größer als 100 sein. Zudem müssen die Variablen normalverteilt sein und ein metrisches Skalenniveau aufweisen (Backhaus et al., 2003). Für die Überprüfung der Modellgültigkeit wurden die globalen Anpassungsmaße Chi-Quadrat-Wert, Comparative Fit Index sowie Root Mean Square Error of Approximation verwendet. Der relativierte Chi-Quadrat-Wert (χ^2/df) sollte kleiner bzw. gleich 2.5 sein (Buch, 2007). Dieser Index wird jedoch kritisch diskutiert, da er durch Faktoren, wie beispielsweise die Stichprobengröße, beeinflusst werden kann und somit schnell über die kritische Grenze steigt (Kreuzpointner, 2010). Einen weiteren Wert für die Güte des Modells liefert der Comparative Fit Index (CFI), der nach Buch (2007), sofern das Modell gültig ist, über .90 liegen sollte. Der letzte Index, der für die Prüfung der Modellgültigkeit betrachtet wurde, ist der Root Mean Square Error of Approximation (RMSEA). Dieser sollte die als kritisch angegebene Grenze von .08 nicht überschreiten (Kline, 2004; zitiert nach Kreuzpointner, 2010).

3.4 Untersuchungsmethode

Die Testungen erfolgten durch die aktuellen Versionen des RLMI-K/J und des AMG-S für Kinder und Jugendliche sowie der überarbeiteten Form des LPS, dem LPS-neu. Das RLMI-K/J wurde am Lehrstuhl von Prof. Lukesch an der Universität Regensburg entwickelt; ebenso wurde das LPS-neu hier überarbeitet. Das AMG-S K-J von Prof. Schmalt ist an der Universität Wuppertal entwickelt worden. Die Komponenten der eben genannten Messinstrumente wurden zusammen mit dem für diese Untersuchung selbst entwickelten Fragebogen der Testmotivation und zusätzlichen Daten in ein Strukturgleichungsmodell integriert, das als Basis für die Hypothesenableitung und die statistische Auswertung diente. Aus diesem Grund wird zunächst das für diese Studie entwickelte Strukturgleichungsmodell erläutert und im Anschluss daran werden die einzelnen Messinstrumente eigens vorgestellt.

3.4.1 Strukturgleichungsmodell

In dieser Studie wurde als Basis für die Ableitung der Hypothesen und für die spätere statistische Auswertung ein Strukturgleichungsmodell verwendet, weshalb an dieser Stelle eine kurze Darstellung erfolgen soll, was unter einem Strukturgleichungsmodell zu verstehen ist; anschließend wird das für diese Studie entwickelte Modell vorgestellt.

Von einem Strukturgleichungsmodell wird gesprochen, wenn die Zusammenhänge der Variablen vorrangig in Gleichungsform dargestellt werden (Buch, 2007). Nach Backhaus und Kollegen (2008) ermöglicht dieses Modell die simultane Überprüfung der Zusammenhänge zwischen latenten Variablen, die sich einer direkten Messung entziehen, und beobachtbaren Variablen. Dabei wird versucht, jede latente Variable mithilfe von Indikatoren (Items), welche als empirische Entsprechung angenommen werden, zu definieren und über diese Indikatorvariablen die latenten Variablen zu messen. Zudem können die latenten Variablen sowohl exogene bzw. unabhängige als auch endogene bzw. abhängige Variablen sein (Buch, 2007). Endogene Variablen können durch die Größen im Modell erklärt werden und untereinander kausale Beziehungen aufweisen (Backhaus et al., 2008). Demgegenüber stehen die exogenen Variablen, die durch das Modell nicht erklärt werden, sondern vorgegeben sind (Backhaus et al., 2008). Folglich untersucht das Strukturgleichungsmodell sowohl den Einfluss exogen latenter Variablen auf endogen latente Variablen als auch die Beziehungen zwischen diesen Variablen und ihren Indikatorvariablen (Buch, 2007). Die mathematische Darstellung der Strukturgleichungsmodelle erfolgt durch lineare Gleichungssysteme, die grafische Darstellung mithilfe eines Pfaddiagrammes (Buch, 2007).

Das Strukturgleichungsmodell, bezogen auf die in dieser Arbeit verwendeten Messinstrumente und zu erfassenden Variablen, lässt sich als Pfaddiagramm (siehe Abbildung 3.1) darstellen. In diesem Pfaddiagramm wurden die beobachtbaren Variablen, also die Leistungsmotive des RLMI-K/J bzw. des AMG-S K-J, die Items der verschiedenen Subskalen der Testmotivation, die Schulnoten und -arten und die einzelnen Subtests des LPS-neu durch Vierecke dargestellt. Die beobachtbaren Variablen werden durch die latenten Variablen beeinflusst, die als Kreise dargestellt sind, wie beispielsweise das Motiv HE des RLMI-K/J durch das Leistungsmotiv HL. Ein Teil der beobachtbaren Variablen geht allerdings auf Messfehler zurück, die aufgrund der Übersichtlichkeit in der Abbildung 3.1 weggelassen wurden. Diese Messfehler beinhalten

weitere nicht berücksichtigte Facetten der latenten Variablen sowie Störfaktoren, die ebenfalls die Messung beeinflussen (Buch, 2007).

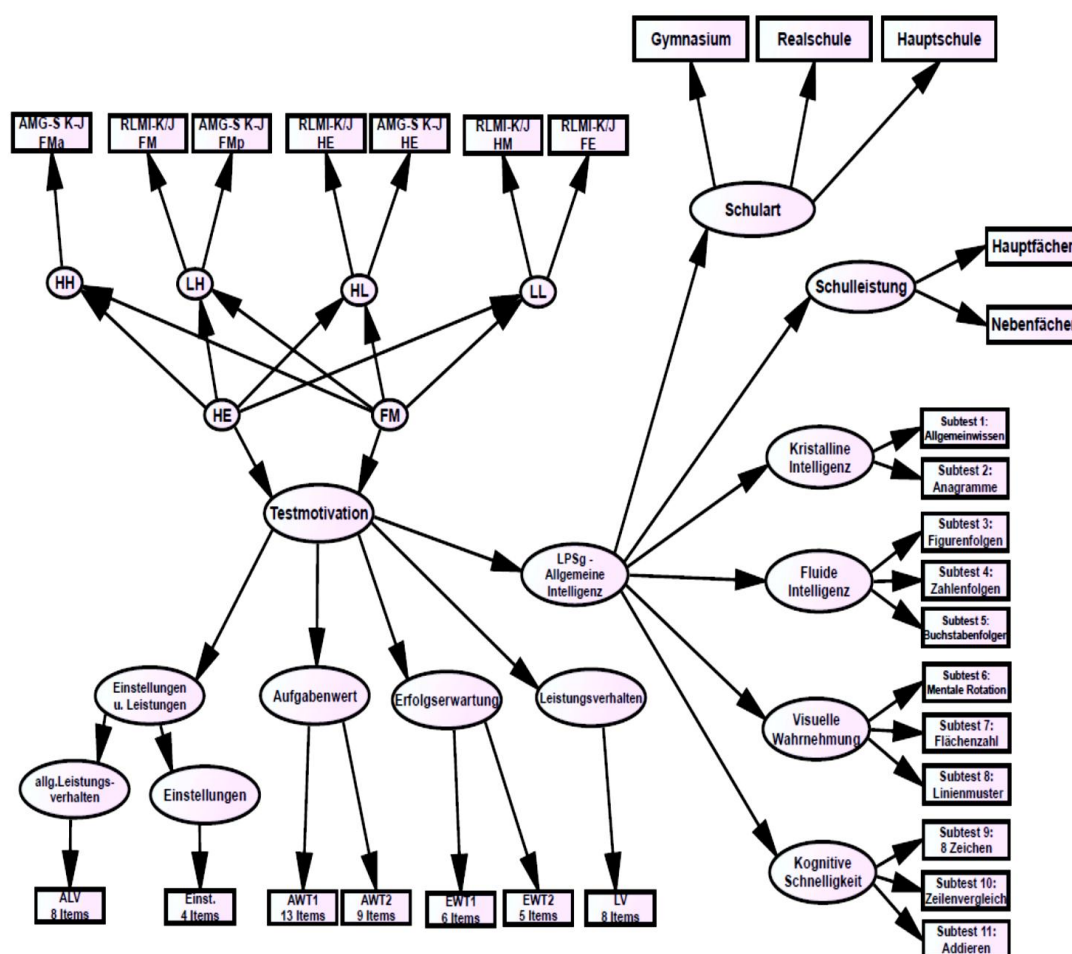


Abbildung 3.1

Strukturmodell dieser Arbeit, dargestellt als Pfaddiagramm

Zusammenfassend lassen sich für das Strukturmodell aus der Abbildung 3.1 folgende Aussagen formulieren: Die Testmotivation ergibt sich aus den Skalen – Einstellungen und Leistungen, Aufgabenwert, Erfolgserwartung und Leistungsverhalten – und wird von den zwei unabhängigen Leistungsmotiven FM und HE der Motive – HL, HH, LH und LL – beeinflusst. Die Testmotivation wiederum übt einen Einfluss auf die allgemeine Intelligenzleistung des LPS-neu aus. Die allgemeine Intelligenzleistung hat einen Einfluss auf die Schulleistungen sowie die Schulart und kann durch die vier Stratum II Komponenten – kristalline Intelligenz, fluide Intelligenz, visuelle Wahrnehmung und kognitive Schnelligkeit – erfasst werden. Welchen angenommenen Einfluss die latenten Variablen auf die beobachtbaren Variablen nehmen und welche Hypothesen daraus für

diese Arbeit abgeleitet werden, kann dem Kapitel 2 entnommen werden. Es wurde jedoch aufgrund der Komplexität und der teilweise nur für grafische Zwecke dargestellten Zusammenhänge darauf verzichtet, das komplette Modell aus Abbildung 3.1 zu überprüfen. Lediglich Teilaspekte wurden mithilfe von Modellrechnungen in AMOS überprüft. Aspekte, die nicht mit Modellrechnungen berechnet wurden, sind mit weiteren statistischen Verfahren getestet worden.

3.4.2 Einverständniserklärung, soziodemographischer Fragebogen und Ergebnismrückmeldung

Das Elternanschreiben (siehe Anhang A-1) umfasst eine kurze Beschreibung der Untersuchung sowie die Einverständniserklärung, die von den Eltern bei Zustimmung der Teilnahme ihres Kindes an der Testung, unterschrieben werden musste.

Mit dem Fragebogen zur Person wurden das Geburtsjahr, das Geschlecht, die besuchte Schulart und Jahrgangsstufe im Schuljahr 2009/2010 (ausgenommen an der Grund- und Mittelschule) sowie die Noten in den Haupt- und Nebenfächern (siehe Anhang A-2) erfasst.

Der Ergebnismrückmeldung an die Schulleiter und an die teilnehmenden Klassen (siehe Anhang A-3) kann die Beschreibung entnommen werden, wie die Ergebnisse des Leistungstests abgerufen werden konnten. Im Anhang A-4 ist die Maske dargestellt, welche die Schüler/-innen in Abhängigkeit von ihrem Versuchspersonencode erhielten. Hier folgte eine kurze Erklärung dazu, was unter IQ zu verstehen sei, welcher Subtest des LPS-neu welche Intelligenzform misst und wie die Schüler/-innen im Vergleich zu ihren Klassenkameraden, zu ihrer Schule und zur Gesamtstichprobe in den verschiedenen Dimensionen des LPS-neu sowie ihrem Gesamtwert abgeschnitten haben.

3.4.3 Das Regensburger Leistungs-Motiv-Inventar für Kinder und Jugendliche (RLMI-K/J)

Mit dem Ziel, die unterschiedlichen Leistungsmotive bei Schüler/-innen erfassen zu können, wurde das Regensburger Leistungs-Motiv-Inventar für Kinder und Jugendliche (RLMI-K/J; Keil, 2009; Sowa, 2009; Schmidbauer, 2008; Lübke, 2007; Peters-Häderle, 2006) in dieser Studie verwendet.

3.4.3.1 Beschreibung

Das RLMI-K/J, das als Weiterentwicklung aus dem RLMI für Erwachsene (RLMI-E) konzipiert wurde, dient der Erfassung der Leistungsmotive für Kinder und Jugendliche im Alter von 12 bis 18 Jahren (Keil, 2009; Sowa, 2009; Schmidbauer, 2008; Lübke, 2007; Peters-Häderle, 2006). Dabei nimmt das RLMI-K/J, genau wie das für Erwachsene auch, eine Unterscheidung der Leistungsmotivation in vier verschiedene Leistungsmotivkomponenten vor: Hoffnung auf Erfolg, Furcht vor Misserfolg, Hoffnung auf Misserfolg und Furcht vor Erfolg (HE, FM, HM, FE). Diese sollen mithilfe der Szenariotechnik, bei welcher leistungsthematische Situationen sowie Antwortaufgaben zu diesen Situationen sprachlich vorformuliert dargestellt sind, erfasst werden.

Im Gegensatz zum RLMI für Erwachsene besteht das RLMI-K/J nicht aus sechs verschiedenen Bereichen, sondern infolge inhaltlicher und testökonomischer Gründe nur aus den beiden Bereichen Ausbildung und Freizeit. Das RLMI-K/J enthält 12 verschiedene Situationen mit vier unterschiedlichen Antwortvorgaben. Jeweils sechs Situationen können dem Bereich Ausbildung und dem Bereich Freizeit zugeordnet werden. Die vier Antwortvorgaben sind so konzipiert, dass sich jeweils eine Vorgabe auf eines der vier Leistungsmotive bezieht. Die Aufgabe der Versuchsperson ist es, je nach persönlicher Einstellung, mithilfe einer fünfstufigen äquidistanten Antwortskala von „Dieser Aussage stimme ich nicht zu“ bis „Dieser Aussage stimme ich sehr zu“ Stellung zu nehmen (Keil, 2009; Sowa, 2009; Schmidbauer, 2008; Lübke, 2007; Peters-Häderle, 2006).

Für das RLMI-K/J wurden zwei Parallelförmigkeiten entwickelt, in welchen die leistungsthematischen Situationen identisch sind. Die beiden Formen unterscheiden sich lediglich in den Statements für die vier Leistungsmotive. Zudem unterscheidet das RLMI-K/J im Vergleich zum RLMI-E nicht zwischen Jungen und Mädchen. In dieser Studie wurde für alle Versuchspersonen ausschließlich die Fragebogenversion A verwendet. Die Bearbeitung dieser Version nahm zwischen 10 und 15 Minuten in Anspruch, wobei es keine Zeitbegrenzung gibt.

3.4.3.2 Testanalyse

Die Vorform des RLMI-K/J wurde an 159 Schüler/-innen in einer Studie von Peters-Häderle (2006) geprüft. Auch diese Form bestand aus 12 Situationen, jedoch mit acht Antwortvorgaben, wobei sich jeweils zwei auf eine der vier Leistungsmotive bezogen. Aufgrund der sich dabei ergebenden sehr guten Homogenitätskoeffizienten der Skalen-

versionen wurden diese Daten für die Entwicklung von Parallelversionen verwendet, wobei Skalen mit vergleichbaren Mittelwerten und Streuungen entstanden sind. Zudem wurde eine abgewandelte Form des RLMI-K/J, für die Entwicklung einer Kurzversion, ebenfalls von Peters-Häderle (2006) an 87 Student/-innen geprüft. Auch hier konnten für alle Bereiche zur Schätzung der internen Konsistenz gute bis sehr gute Werte von Cronbach's Alpha (Cronbach's α) erreicht werden.

Wie die diesbezüglichen Werte in dieser Studie ausgefallen sind, kann für die beiden Bereiche Ausbildung und Freizeit, aufgeteilt nach den vier Leistungsmotiven, der Tabelle 3.5 entnommen werden. Zunächst mussten jedoch die Daten bereinigt werden, da einige Versuchspersonen zu viele fehlende Angaben hatten. Bei Skalen, bei denen ein Item zur Vervollständigung fehlte, wurde es durch Mittelwerte ersetzt. Dementsprechend konnten 348 Versuchspersonen in der Berechnung berücksichtigt werden. Insgesamt wurden 29 einzelne Items durch Mittelwerte ersetzt.

Tabelle 3.5

Überblick über die Maße der internen Konsistenz (Cronbach's α)

	HE	FM	FE	HM
Ausbildung	.59	.63	.75	.76
Freizeit	.56	.64	.70	.68
Gesamt	.74	.77	.84	.83

Es zeigen sich für die Gesamtskalen ausreichende bis hohe Werte von .74 bis .84. Insgesamt weist der Bereich Ausbildung etwas höhere Werte für die internen Konsistenzen auf als der Bereich Freizeit.

Wie die verschiedenen Gesamtskalenwerte miteinander zusammenhängen, ist in Tabelle 3.6 dargestellt. Die Berechnungen der bivariaten Korrelationen erfolgten nach Bravais-Pearson. Die drei Skalen FM, FE und HM, die im Gegensatz zur Skala HE Misserfolgskognitionen beinhalten, stehen in einem mittleren bis sehr hohen positiven Zusammenhang zwischen .49 und .80. Die Skala HE weist hingegen schwächere und teilweise signifikant negative Zusammenhänge zu den anderen drei Skalen, zwischen -.06 und -.17, auf.

Tabelle 3.6

Interkorrelation der Skalen des RLMI-K/J

	HE	FM	FE	HM
HE	-	-.11*	-.06	-.17**
FM		-	.49**	.50**
FE			-	.80**
HM				-

Anmerkungen. **) Die Korrelation ist auf dem .01 Niveau signifikant; *) Die Korrelation ist auf dem .05 Niveau signifikant

Aus der Tabelle B-1 (siehe Anhang) werden die Itemzahlen, die Mittelwerte und die Standardabweichungen der vier Leistungsmotive des RLMI-K/J ersichtlich, getrennt für die Bereiche Ausbildung und Freizeit und für beide zusammen sowie die Schiefe und Kurtosis (Exzess) der Verteilungen für die berücksichtigte Stichprobe. Die beiden Leistungsmotive HE und FM weisen in allen Bereichen eine geringe Schiefe von $<.30$ auf. Mithilfe der Schiefe kann überprüft werden, inwieweit die vorliegende Form einer Normalverteilung gleicht. Bei einer geringen Schiefe wird eine annähernde Symmetrie der Verteilung angenommen. Bei den anderen beiden Leistungsmotiven fällt vor allem die hohe Schiefe von 1.00 der Skala HM im Bereich Ausbildung auf. Auch die weiteren Bereiche dieser beiden Skalen weisen mit Werten für die Schiefe von über .59 auf jeweils linksschiefe Verteilungen hin, dementsprechend werden hier kleinere Zustimmungswerte bevorzugt. Auch für die Kurtosis, die die Wölbung einer Verteilung angibt, stellen sich bei allen Skalen nur geringfügige Abweichungen von der Normalverteilung (siehe Tabelle B-1; Anhang B) heraus. Die höchsten Werte zeigen sich hier bei der Leistungsmotivkomponente HM in der Ausbildung mit .90, was auf eine etwas steilgipfligere Verteilung hinweist.

3.4.4 Das Leistungsmotivgitter für Kinder und Jugendliche (Kurzversion)(AMG-S K-J)

Auch das Leistungsmotivgitter für Kinder und Jugendliche in der Kurzversion (AMG-S K-J; Schmalt 1999; 2005) wurde für die Erhebung des Leistungsmotivs bei Schüler/-innen in dieser Studie verwendet.

3.4.4.1 Beschreibung

Das von Schmalt (1976; 1999) entwickelte LM-Gitter ist dem Verfahren des TAT ähnlich. Hierbei erhalten die Probanden Bilder und darunter Aussagen zu den Bildern. Es handelt sich somit um ein semi-projektives Verfahren. Die Aufgabe der Probanden besteht darin, die Aussagen zu wählen, die für sie am besten zu den Bildern passen. Es wird dabei angenommen, dass die Auswahl durch die Motive einer Person beeinflusst wird (Langens, Schmalt & Sokolowski, 2005).

Das in dieser Studie verwendete LM-Gitter für Kinder und Jugendliche in der Kurzversion (AMG-S K-J; Schmalt, 1999; 2005) ist eine Weiterentwicklung des LM-Gitters für Erwachsene und wurde für Kinder und Jugendliche im Alter zwischen 9 bis 16 Jahren konzipiert. Die Langversion sowohl des LM-Gitters für Erwachsene als auch die für Kinder und Jugendliche besteht aus 18 nicht ganz deutlich grafisch dargestellten Leistungssituationen. Unter jedem der 18 Bilder stehen jeweils 18 gleiche Aussagen, die den Inhaltskategorien des TAT entlehnt sind (Brunstein & Heckhausen, 2006). Die in dieser Studie verwendete Kurzversion besteht hingegen aus lediglich sechs grafischen Situationen, in denen die Leistungsmotivation eine Rolle spielen kann. In den verwendeten Bildern sind die Akteure geschlechtsneutral dargestellt und ein erfolgreicher bzw. erfolgloser Ausgang der Situation ist gleich wahrscheinlich. Dabei können die Bilder den Kategorien – Singen im Chor, Schwimmen, Klavier spielen, Schule, Seil klettern und Zimmern – zugeordnet werden (Schmalt, 2005). Unterhalb der Bilder sind zehn leistungsthematische Aussagen aufgelistet, wobei jeweils drei der Aussagen, entsprechend der Aufteilung des Leistungsmotivs im LM-Gitters für Erwachsene, einer Motivkomponente zugeordnet werden kann: Hoffnung auf Erfolg, Furcht vor Misserfolg aktiv und Furcht vor Misserfolg passiv (HE, FMa, FMp). Das Item – „Er fühlt sich wohl dabei.“ –, das unter jedem Bild als Einstieg dient, bezeichnet Schmalt (2005) als Füller und kann keiner der drei Motivkomponenten zugeordnet werden. Deshalb bleibt es bei den Skalenberechnungen unberücksichtigt. Auch die restlichen neun Aussagen sind bei jedem Bild identisch, nur die Reihenfolge der Items unterscheidet sich von Bild zu Bild. Der Proband hat nun die Aufgabe, die Aussagen anzukreuzen, die seiner Einstellung zufolge auf die dargestellte Person im Bild passen (Brunstein & Heckhausen, 2006). Dabei kann er so viele Aussagen ankreuzen wie er möchte. Die Aussagen, die seiner Meinung nach nicht zutreffen, sollen mit einer „0“ gekennzeichnet werden. Durch dieses vorgegebene Antwortschema soll sowohl die Objektivität als auch die implizite Er-

hebung des Leistungsmotivs gewährleistet sein. Zudem wurden in dieser Studie zwei verschiedene Versionen getrennt für Mädchen und Jungen verwendet. Diese unterscheiden sich lediglich im Anfangswort der Aussagen – „Sie“ für Mädchen und „Er“ für Jungen – und die restlichen Aussagen sowie Bilder sind identisch. Die Bearbeitung der Kurzversion dauert zwischen 10 und 15 Minuten, wobei es auch hier keine Zeitbegrenzung gibt.

3.4.4.2 Testanalyse

Die Kurzform des AMG für Kinder und Jugendliche wurde an jeweils 285 und 153 Schüler/-innen in zwei Studien von Schmalt (2005) geprüft. Die in diesen Studien verwendete Form des AMG-S entspricht auch der in dieser Arbeit angewandten Version. Für die Gesamtskalen der drei Leistungsmotivkomponenten zeigten sich in der ersten Studie etwas höhere Werte von .77 bis .81 für die interne Konsistenz von Cronbach's α . Aber auch in der zweiten von Schmalt (2005) durchgeführten Studie konnten für alle Bereiche zur Schätzung der internen Konsistenz gute bis sehr gute Werte von .76 bis .79 für Cronbach's α erreicht werden.

Wie die diesbezüglichen Werte in dieser Studie ausgefallen sind, kann für die verschiedenen Leistungsmotivkomponenten der Tabelle 3.7 entnommen werden. Zunächst mussten jedoch die Daten bereinigt werden, da einige Versuchspersonen zu viele fehlende Angaben hatten. Dementsprechend konnten 322 Versuchspersonen in der Berechnung berücksichtigt werden. Items wurden nicht durch Mittelwerte ersetzt, da alle Versuchspersonen, bei denen Daten fehlten, mindestens ein Bild und dessen Aussagen nicht bearbeitet hatten.

Tabelle 3.7

Überblick über die Maße der internen Konsistenz (Cronbach's α)

	HE	FMa	FMp
Gesamt	.79	.83	.69

Demnach zeigt sich im Vergleich zu den von Schmalt (2005) durchgeführten Studien ein etwas geringerer Wert der internen Konsistenz für die FMp. Die beiden anderen Werte entsprechen den von Schmalt (2005) berechneten internen Konsistenzen.

In der Tabelle 3.8 sind die Zusammenhänge zwischen den verschiedenen Gesamtskalenwerten dargestellt. Die Berechnungen der bivariaten Korrelationen erfolgten

hier ebenfalls nach Bravais-Pearson. Auch für die Interkorrelationen der Skalen sind in den Studien von Schmalt (2005) Werte angegeben. So korreliert beispielsweise die HE mit der FMa zwischen .18 und .24 signifikant, wobei sich in dieser Studie mit $r = .47$ ein deutlich höherer Zusammenhang zeigt. Für die anderen beiden Zusammenhänge zwischen FMa und FMp bzw. zwischen HE und FMp zeigen sich in den Studien zumindest annähernd ähnliche Werte von $r = .21$ ($p < .01$) bzw. $r = .03$. Entsprechend der Theorie korreliert die FMa signifikant positiv mit der HE und der FMp, da sie sowohl in der Hoffnung auf Erfolg als auch in der Furcht vor Misserfolg hohe Werte aufweist. Hingegen wäre ein negativer Zusammenhang zwischen HE und FMp zu erwarten gewesen.

Tabelle 3.8

Interkorrelation der Skalen des AMG-S K-J

	HE	FMa	FMp
HE	-	.47**	.03
FMa		-	.25**
FMp			-

Anmerkung. **) Die Korrelation ist auf dem .01 Niveau signifikant

Aus der Tabelle B-2 (siehe Anhang) werden die Itemzahlen, die Mittelwerte und die Standardabweichungen der Motive des AMG-S K-J sowie die Schiefe und Kurtosis (Exzess) der Verteilungen für die berücksichtigte Stichprobe ersichtlich. Die Skala HE weist eine sehr geringe Schiefe von $< .30$ auf. Die beiden anderen Skalen weisen eine mittlere Schiefe auf, weshalb bei allen Skalen von einer annähernden Symmetrie der Verteilung auszugehen ist. Auch für die Kurtosis weisen alle Skalen nur geringfügige Abweichungen von der Normalverteilung auf (siehe Tabelle B-2; Anhang B). Die höchsten Werte zeigen sich hier bei der Leistungsmotivkomponente FMp mit .87, was auf eine etwas steilgipfligere Verteilung hinweist.

3.4.5 Fragebogen zur Erfassung der Testmotivation

Da in dieser Arbeit davon ausgegangen wird, dass der Zusammenhang zwischen Leistungsmotiv und Leistungen in einem Leistungstest über die Testmotivation vermittelt wird, wurde ein Fragebogen entwickelt, der die Motivation in diesem Test zu erfassen versucht.

3.4.5.1 Beschreibung

Es gibt einige Fragebögen, die die Messung von Testmotivation beinhalten, doch häufig besteht der Schwerpunkt auf der Erfassung der Leistungsmotivation. Zudem sollte ein Fragebogen entwickelt werden, welcher auf der Erwartungs-Wert-Theorie von Eccles und Kollegen (1983) basiert. Aus diesen beiden Gründen wurde ein Fragebogen zur Erfassung der Testmotivation entworfen, der sich an verschiedene, bereits bestehende Fragebögen anlehnt bzw. Items von diesen beinhaltet und zusätzlich auf selbst entworfene Items zurückgreift.

Nach der Erwartungs-Wert-Theorie von Eccles und Kollegen (1983) haben verschiedene Komponenten der Leistungsmotivation Einfluss auf die Testmotivation, weshalb der entwickelte Fragebogen aus insgesamt fünf verschiedenen Skalen besteht: allgemeines Leistungsverhalten (ALV) und grundsätzliche Einstellungen (Einst.), zusammengefasst zu Einstellungen und Leistungen (EL), Leistungsverhalten im Test (LV), Aufgabenwert (AW) und Erfolgserwartung (EW). Zusammen ergeben sie den Gesamtwert der Testmotivation (siehe Kapitel 1.3.2; siehe Anhang A-5). Bei der Entwicklung des Fragebogens wurde sich für eine einheitliche fünfstufige Antwortskala der einzelnen Items entschieden. Auch die aus bereits bestehenden Fragebögen stammenden Items wurden an diese Antwortskala angepasst.

Die Skala ALV umfasst die Items, die sich mit der Erwartung und Wertigkeit einer Aufgabe befassen, da Wigfield und Eccles (2000) annehmen, dass diese die Leistungswahl, die Fähigkeit sowie die Anstrengung und die Ausdauer des Leistungsverhaltens direkt beeinflussen. Im Fragebogen zur Leistungsmotivation für Schüler der 7. bis 13. Klasse (FLM 7-13) von Petermann und Winkel (2007) werden diese Komponenten erfasst. Aus diesem Grund wurden die ersten Items des Fragebogens zur Testmotivation „vor dem Test“ (siehe Anhang A-5; vA1 bis vA7) für die Skala ALV direkt aus dem FLM 7-13 (Petermann & Winkel, 2007, S. 26) übernommen. Auch das Item v1 (siehe Anhang A-5) aus dem Fragebogen von Eklöf (2008, S. 13) beinhaltet die Anstrengungsbereitschaft. Es wurde deshalb ebenfalls der Skala ALV zugeordnet.

Die nächsten vier Items aus dem Fragebogen „vor dem Test“ (siehe Anhang A-5; vB1 bis vB4) erfassen die grundsätzlichen Einstellungen – individuelle Ziele, bisherige Erfolgs- und Misserfolgserfahrungen und die Einstellungen von Bezugspersonen –, wobei diese Einfluss auf die Testmotivation ausüben. Die ersten beiden Items der Skala

wurden selbst entwickelt, die beiden letzten stammen sinngemäß aus dem Fragebogen von Eklöf (2008). Zusammen bilden die beiden eben genannten Skalen, ALV und Einst., die Skala „Einstellungen und Leistungen“, da sie sowohl grundsätzliche Einstellungen zu Leistungsthemen als auch allgemeine Einstellungen zum Inhalt haben.

Die nächste Skala AW wurde aufgrund der Wichtigkeit für die Erwartungs-Wert-Theorie (Eccles et al., 1983) sowohl vor (AWT1) als auch nach dem Test (AWT2) erfragt (siehe Anhang A-5; Items vC1 bis vC11, v2, v3 und nA1 bis nA8, n1). Dabei gehören nach Eccles und Kollegen (1983) vier Komponenten zu dieser Skala: Wichtigkeitskomponente, intrinsischer Wert, Nützlichkeitskomponente und Kosten. Die ersten drei Items aus dem Fragebogen „vor dem Test“ umfassen die Wichtigkeit einer Aufgabe und stammen sinngemäß aus verschiedenen Fragebögen (Pohlmann et al., 2005, S. 130; Arvey et al., 1990, S. 714; Guay, Vallerand & Blanchard, 2000, S. 184). Der intrinsische Wert wird durch die nächsten drei Items (siehe Anhang A-5, vC4 bis vC6) erfasst. Dabei stammen die ersten beiden Aussagen, in teilweise leicht abgeänderter Form, aus dem Fragebogen zur Erfassung aktueller Motivation in Lern- und Leistungssituationen (FAM) von Rheinberg, Vollmeyer und Burns (2001, S. 17); das letzte Item wurde selbst entwickelt. Die Aussagen vC7 und vC8 aus dem Fragebogen „vor dem Test“ können der Komponente Nützlichkeit zugeordnet werden. Dabei wurde das erste Item sinngemäß aus dem Fragebogen von Wigfield und Eccles (2000, S. 70) entnommen und das letzte aus The Situational Motivation Scale (SIMS) von Guay et al. (2000, S. 184). Die letzten drei Items gehören zur Komponente Kosten, wobei zwei Items selbst entwickelt wurden und das Item vC11 aus einem Fragebogen von Cole (2007, S. 85) stammt. Die Items v2 und v3, welche ebenfalls zur Skala AWT1 gehören, wurden dem Fragebogen von Eklöf (2008, S. 13) entnommen. Auch die Skala AWT2 besteht jeweils aus zwei Items zur Wichtigkeit, zum intrinsischen Wert, zur Nützlichkeit und zu den Kosten. Die ersten beiden Items stammen aus den Fragebögen von Pohlmann et al. (2000, S. 130) und von Arvey (1990, S. 714). Die Items nB3 und nB4, zugehörig zum intrinsischen Wert, wurden dem FAM von Rheinberg et al. (2001, S. 17) entnommen. Die Aussagen zur Nützlichkeit stammen aus dem Fragebogen von Wigfield und Eccles (2000, S. 70). Die Kosten werden durch ein sinngemäßes Item von Cole (2007, S. 85) und durch ein selbst entwickeltes Item erfasst. Das letzte Item n1, welches ebenfalls zum AWT2 gehört, wurde dem Fragebogen von Eklöf (2008, S. 13) entnommen.

Auch die Skala EW wurde aufgrund ihrer Wichtigkeit sowohl vor (EWT1) als auch nach dem Test (EWT2) erhoben (siehe Anhang A-5; Item vD1 bis vD5, v4 und nC1 bis

nC4, n2). Gemäß der Modellannahmen wird die Erfolgserwartung sowohl vom Begabungsselbstkonzept, vom Fähigkeitsselbstkonzept als auch von der wahrgenommenen Aufgabenschwierigkeit direkt bestimmt (Pohlmann et al., 2005; Wigfield & Eccles, 2000). Demnach verteilen sich diese Komponenten auf die zu dieser Skala gehörenden Items, wobei lediglich vD3 aus einem bereits bestehenden Fragebogen (Pohlmann et al., 2005, S. 130) stammt und die restlichen selbst entwickelt wurden. Auch bei den Items zur EWT2 wurden lediglich die Items nC1 und n2 aus einem Fragebogen sinngemäß entnommen (Pohlmann et al., 2005, S. 130; Sundre 2007, S. 15). Die anderen Aussagen sind selbst entwickelt und beinhalten das Fähigkeitsselbstkonzept und die Aufgabenschwierigkeit.

Die letzte Skala des Fragebogens zur Testmotivation LV (siehe Anhang A-5; Items nA1 bis nA7, n3) wurde nur nach dem Test erfragt und beinhaltet dieselben Komponenten wie die Skala ALV. Eine Unterscheidung wurde deshalb vorgenommen, weil sich die Skala ALV auf das allgemeine Leistungsverhalten bezieht, nicht zwingend abhängig von Testsituationen, und die Skala LV nur auf die Motivation des vorangegangenen Tests Bezug nimmt. Dabei wurden die ersten vier Items, welche die Anstrengung und Ausdauer thematisieren, teilweise sinngemäß dem Fragebogen The Student Opinion Scale (SOS; Sundre, 2007, S. 15) entnommen. Ebenfalls zur Komponente Ausdauer gehört das Item nA5 und stammt aus dem FLM 7-13 von Petermann und Winkel (2007, S. 26). Die letzten beiden Aussagen zur Leistungswahl wurden selbst entwickelt. Das Item n3 schließlich gehört zum Fragebogen von Rauner und Kollegen (2009, S. 225).

3.4.5.2 Testanalyse

Bevor die eigentliche Analyse des Fragebogens zur Testmotivation vorgenommen werden konnte, mussten die negativ formulierten Items umkodiert werden (siehe Anhang A-5; betroffene Items sind mit einem „u“ gekennzeichnet). Zudem wurden die Daten nachbearbeitet, da viele Versuchspersonen ganze Skalen nicht bearbeitet hatten. Bei Versuchspersonen, bei denen nur jeweils ein Item zur Vervollständigung einer Skala fehlte, wurden diese durch Mittelwerte ersetzt. Somit konnten für die Testmotivation 334 Versuchspersonen berücksichtigt werden und es wurden insgesamt 8 Items durch Mittelwerte ersetzt.

Im Anschluss daran wurden die internen Konsistenzen für Cronbach's α der einzelnen Skalen und die Split-Half-Koeffizienten nach der Spearman-Brown-Formel berech-

net. Die beiden Skalen AW und EW werden zu einer Gesamtskala, welche sich aus den Werten vor und nach dem Test ergeben, addiert. Die Ergebnisse können der Tabelle 3.9 entnommen werden.

Tabelle 3.9

Überblick über die Maße der internen Konsistenz (Cronbach's α) und Split-Half-Koeffizienten nach Spearman-Brown-Formel

	AW	rtt ¹	EW	rtt ¹	ALV	rtt ¹	Einst.	rtt ¹	LV	rtt ¹
Vor dem Test	.87	.85	.70	.69	.63	.53	.33	.18	-	-
Nach dem Test	.83	.85	.53	.54	-	-	-	-	.77	.72
Gesamt	.91	.89	.70	.74	.68	.65	.68	.65	.77	.72

Anmerkungen. ALV und Einst. ergeben zusammen die Skala EL; rtt¹ = korrigierter Split- Half-Koeffizient nach Spearman-Brown-Formel; ¹Teilung in zwei gleichgroße Skalen

Es zeigen sich für die Skalen akzeptable bis sehr gute Werte, wobei lediglich die Skala Einst. mit sehr geringen Werten von .33 für die interne Konsistenz und einem Split-Half-Koeffizienten nach Spearman-Brown-Formel von .18 auffällt. Diese Skala bildet jedoch zusammen mit der Skala ALV die Skala EL und wird deshalb weiter berücksichtigt, da sich für die Gesamtskala ein guter Wert für die interne Konsistenz von .68 ergibt. Darüber hinaus können aus der Tabelle B-3 (siehe Anhang) die Itemzahlen, die Mittelwerte und die Standardabweichungen sowie die Schiefe, die Kurtosis (Exzess) als auch die mittleren Trennschärfen für die einzelnen Skalen des Fragebogens der Testmotivation entnommen werden. Die mittleren Trennschärfen wurden mithilfe der Fisher Inverse berechnet. Die Schiefe ist außer bei der Skala LV mit -.48 als gering einzuschätzen. Die Kurtosis weist vor allem bei der Skala Aufgabenwert mit -.98 auf eine flache Verteilung hin. Es kann jedoch bei allen Skalen von einer annähernden Symmetrie der Verteilungen ausgegangen werden. Auch bei der mittleren Trennschärfe fällt die Skala Einst. wieder mit einem sehr geringen Wert von .18 auf, wobei diese zusammen mit der Skala ALV dennoch einen akzeptablen Wert aufweist. Alle anderen Skalen weisen respektable Werte bei diesem Koeffizienten auf. Bei genauerer Betrachtung der Itemkennwerte des Fragebogens zur Testmotivation (Tabelle B-4; Anhang B) fällt wieder die Skala Einst. auf, die mit nur einem Item eine mittlere Trennschärfe aufweisen kann. Die Trennschärfen geben hier durch die Korrelation des Items mit dem Gesamtestwert an, wie gut das Item das Gesamtergebnis des Fragebogens repräsentiert. Zudem wurden in der Tabelle B-4 (siehe Anhang) alle Items gekennzeichnet, die keine ausrei-

chende Trennschärfe aufweisen und für jedes Item wurden die Mittelwerte und Standardabweichungen berechnet.

Betrachtet man die bivariaten Korrelationen nach Bravais-Pearson aus Tabelle 3.10, so weisen die Skalen des Aufgabenwertes die höchsten Korrelationen mit der Testmotivation auf und korrelieren sehr hoch mit der Skala EL. Die Erfolgserwartung vor dem Test zeigt den geringsten Zusammenhang mit der Testmotivation. Darüber hinaus korrelieren die Skalen insgesamt alle signifikant untereinander, wobei sich die höchsten Zusammenhänge zwischen Aufgabenwert vor dem Test (AWT1) und Einstellungen und Leistungen (EL) zeigen (siehe Tabelle 3.10).

Tabelle 3.10

Interkorrelation der Skalen des Fragebogens zur Erfassung der Testmotivation

	TMg	AWT1	AWT2	AW	Einst.	ALV	EL	LV	EWT1	EWT2	EW
TMg	-	.84	.84	.93	.56	.64	.71	.67	.43	.58	.60
AWT1		-	.65	.93	.51	.49	.58	.40	.20	.27	.28
AWT2			-	.88	.38	.39	.45	.56	.20	.46	.39
AW				-	.50	.50	.57	.52	.22	.39	.36
Einst.					-	.45	.75	.18	.22	.15	.23
ALV						-	.93	.25	.34	.21	.34
EL							-	.26	.34	.22	.35
LV								-	.18	.46	.38
EWT1									-	.34	.84
EWT2										-	.80
EW											-

Anmerkungen. TMg = gesamte Testmotivation; alle Korrelationen sind signifikant auf dem Niveau von .01

Zudem korreliert die TM vor dem Test (TMv), die sich aus den Skalen Einst., ALV, AWT1 und EWT1 ergibt, etwas höher mit der TMg ($r = .91$) als die TM nach dem Test (TMn) ($r = .87$), bestehend aus LV, AWT2 und EWT2. Zusammen korrelieren diese beiden mit .58.

3.4.5.3 Konfirmatorische Faktorenanalyse

Nach Abschluss der Datenerhebung und Eingabe der Rohdaten in das Programm wird im Rahmen der konfirmatorischen Faktorenanalyse als Nächstes nach Ausreißern ge-

sucht. Für die Suche wurde die Mahalanobis Distanz verwendet. Sie bezeichnet den Abstand eines n -dimensionalen Punktes vom n -dimensionalen Mittelwert (Centroid):

$$d^2 = (X^n - \mu^n)' \Sigma^{-1} (X^n - \mu^n)$$

(Bortz, 1999; zitiert nach Kreuzpointner, 2010).

Dementsprechend wurde nach Versuchspersonen gesucht, deren Testmotivation erheblich von der durchschnittlichen Testmotivation abweicht. Hierbei fielen zwei Versuchspersonen mit einem Abstand von 31.76 und 28.99 auf, bei einem durchschnittlichen Abstand von 6.98 und einer Standardabweichung von 4.95. Diese beiden Probanden wurden aufgrund auffällig hoher bzw. niedriger Rohwerte aus den weiteren Berechnungen ausgeschlossen. Für weitere sechs Versuchspersonen war der Abstand zwar signifikant, aber sie zeigten keine auffälligen Rohwerte, sodass diese weiter berücksichtigt wurden und somit insgesamt 332 Versuchspersonen in die weiteren Berechnungen einfließen.

Als nächster Schritt der konfirmatorischen Faktorenanalyse folgt nochmals eine zusammenfassende Darstellung der Interkorrelationen, Kovarianzen, Mittelwerte, Standardabweichungen, Schiefe- und Kurtosisparameter der verschiedenen Skalen des Fragebogens zur Testmotivation. Wie der Tabelle 3.11 zu entnehmen ist (siehe auch Kapitel 3.3.4.2), deuten die Werte von Schiefe und Kurtosis (Exzess) für die manifesten Variablen, also für die Skalen, auf eine Normalverteilung des vorliegenden Datensatzes hin. Zudem zeigt die Korrelation zwischen den Skalen mit dem höchsten Wert von .69 zwischen AWT1 und AWT2, dass keine Multikollinearität bezüglich dieser Skalen des Testmotivationsfragebogens besteht. Die quadrierte multiple Korrelation zwischen jeder der sieben Skalen und den sechs anderen ergibt einen R^2_{smc} – Wert von .60 für die Skala AWT2. In den Berechnungen war das der höchste Wert, der aber deutlich unter der Grenze von .90 liegt (Kline, 2004 zitiert nach Kreuzpointner, 2010).

Tabelle 3.11

Interkorrelationen, Kovarianzen, Mittelwerte, Standardabweichungen, Schiefe- und Kurtosisparameter der verschiedenen Skalen der Testmotivation

	ALV	Einst.	AWT1	AWT2	LV	EWT1	EWT2
ALV	19.98	5.11	20.89	12.70	6.42	5.29	3.09
Einst.	.45	6.50	12.03	7.35	2.72	2.06	1.28
AWT1	.50	.50	88.50	45.99	22.38	7.08	8.12
AWT2	.40	.40	.69	51.15	22.96	4.88	10.62
LV	.25	.19	.42	.56	32.63	3.41	8.52
EWT1	.34	.23	.21	.19	.17	12.46	3.93
EWT2	.22	.16	.27	.47	.47	.35	10.20
Mittelwert	26.79	11.44	38.74	24.72	28.42	20.71	15.64
SD	4.47	2.55	9.41	7.15	5.71	3.53	3.19
Schiefe	-.27	.01	.04	.11	-.48	-.16	-.11
Kurtosis	-.64	-.73	-.95	-.99	-.68	-.14	-.62

Anmerkungen. die grau markierte Diagonale zeigt die Varianzen an; über der Diagonalen stehen Kovarianzen, unter der Diagonalen Korrelationen; SD = Standardabweichung; für $n = 332$ sind alle Korrelationen signifikant auf dem Niveau von .01

Im Rahmen der konfirmatorischen Faktorenanalyse werden auch die Reliabilitäten der manifesten Variablen betrachtet. Dabei kann die interne Konsistenz der Tabelle 3.9 entnommen werden und die Split-Half-Reliabilität der Tabelle B-3 im Anhang. Mit Ausnahme der Skala Einst. ($\alpha = .33$; $r_{tt} = .18$) liegen alle Skalen im adäquaten Bereich.

Als Nächstes folgt die mathematische Bestimmung des Modells zur Testmotivation. Die Struktur der konfirmatorischen Faktorenanalyse und die dabei vermuteten Zusammenhänge sind in der Abbildung 2.2 ersichtlich, wobei die Fehler- und Residualvariablen der Abbildung 3.2 sowie der Tabelle 3.12 entnommen werden können. Für die Lösung dieses Modells müssen mindestens so viele Parameter, die zu schätzen sind, aus den Daten bestimmt werden (Kreuzpointner, 2010). Die Daten liefern 35 Werte und geschätzt werden insgesamt 24. Somit beträgt die Zahl der Freiheitsgrade für das Modell 11, weshalb es identifizierbar ist und mithilfe von AMOS berechnet werden kann. Das mit AMOS berechnete Strukturgleichungsmodell zur Testmotivation für die Stichprobe von 332 Versuchspersonen ist in Abbildung 3.2 dargestellt.

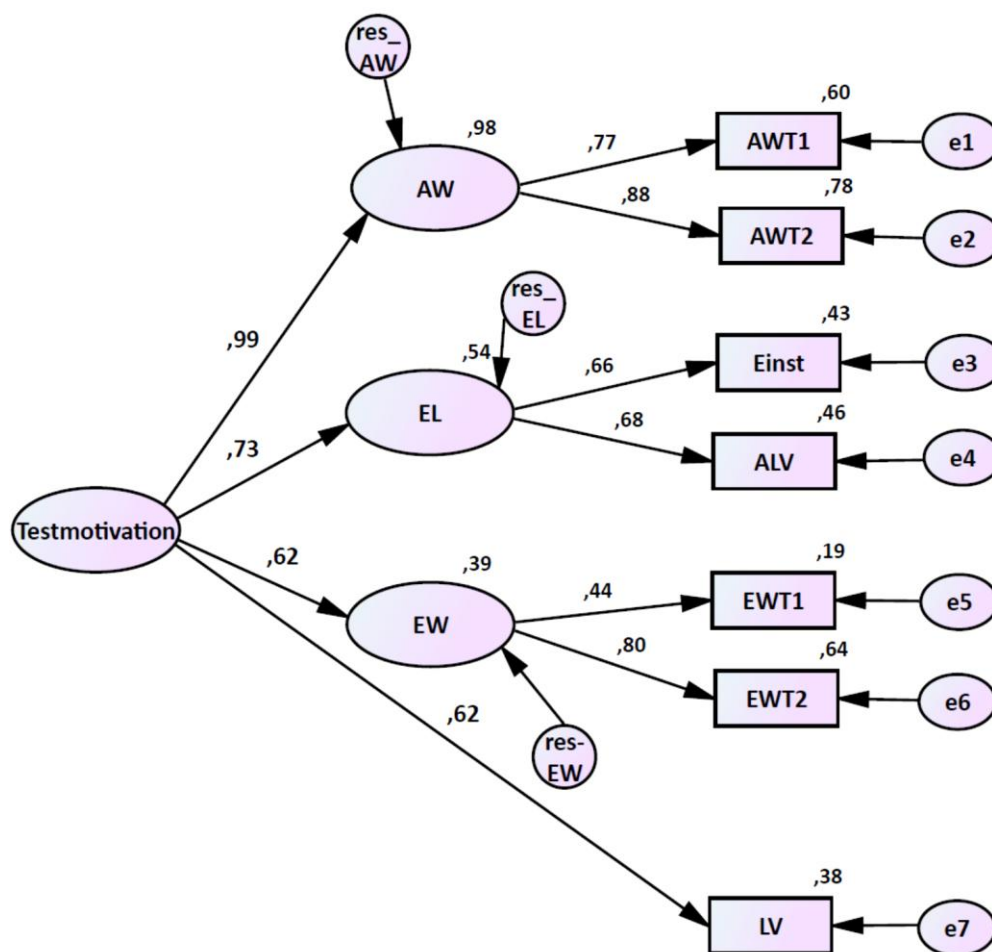


Abbildung 3.2

Modell der Testmotivation

Im Anschluss wird die Güte des Modells mithilfe verschiedener Indizes bewertet. Dabei gibt AMOS eine Reihe von Indizes an. In dieser Arbeit wurde sich jedoch auf drei beschränkt. Der Abbildung 3.2 können zunächst die Korrelationen zwischen den verschiedenen Skalen untereinander und mit der Testmotivation entnommen werden. Dabei fällt die überhöhte Korrelation zwischen der Testmotivation und der Skala AW von .99 besonders auf. Die anderen Korrelationen sind unauffällig. Für das in Abbildung 3.2 formulierte Modell ergibt sich ein χ^2 von 102.62 mit $df = 11$ und $p = .00$. Dieser Index wird jedoch in Strukturgleichungsmodellen kritisch diskutiert, da er durch Faktoren wie beispielsweise die Stichprobengröße beeinflusst werden kann (Kreuzpointner, 2010). Aus diesem Grund sollten zusätzliche Indizes betrachtet werden. Einen weiteren Wert für die Güte des Modells liefert der CFI, welcher nach Kline, sofern das Modell gültig ist, über .90 liegen sollte (Buch, 2007). Für das in Abbildung 3.2 formulierte Modell ergibt sich jedoch lediglich ein CFI von .873, was knapp ist, aber gegen eine Modellgültigkeit spricht. Der letzte Index, der für die Prüfung der Modellgültigkeit betrachtet wird, ist

der RMSEA. Für das in Abbildung 3.2 formulierte Modell ergibt sich ein Wert von .159, der mit 90 % Wahrscheinlichkeit im Intervall [.131; .187] bleibt. Alle drei Werte überschreiten die als kritisch angegebenen Grenzen von .08 und [.05; .10] (Kline, 2004 zitiert nach Kreuzpointner, 2010). Auch dieser Fit-Index deutet daraufhin, dass das Modell so nicht gültig ist.

Als nächster Schritt der konfirmatorischen Faktorenanalyse werden die Parameterschätzungen betrachtet und beurteilt. Die unstandardisierten Parameter und die Standardfehler sowie die Residualvarianz bzw. Fehlervarianz, jeweils unstandardisiert, können der Tabelle 3.12 entnommen werden.

Tabelle 3.12

Unstandardisierte Regressionsgewichte, Standardfehler (in Klammern) und Residualvarianz bzw. Fehlervarianz (unstandardisiert) des Modells

	TM	AW	EL	EW	Residualvarianz
TM					1.52 (0.36)
AW	5.82 (0.73)				1.04 (5.03)
EL	1				1.29 (0.35)
EW	1.30 (0.22)				1.43 (0.45)
LV	2.85 (0.43)				20.17 (2.02)
AWT1		1			35.53 (4.18)
AWT2		0.87 (0.07)			11.11 (2.53)
Einst.			1		3.67 (0.44)
ALV			1.81 (0.23)		10.69 (1.37)
EWT1				1	10.07 (0.91)
EWT2				1.67 (0.36)	3.64 (1.32)

Bei Betrachtung der unstandardisierten Regressionsgewichte fällt die Skala AW mit einem Wert von 5.82 besonders auf, wohingegen alle anderen Werte mindestens unter drei bleiben. Dies könnte auch ein Grund für die überhöhte Korrelation zwischen dieser Skala und der Testmotivation sein. Alle Residualvarianzen sind positiv, wobei auch bei der Fehlervarianz vor allem die Skala AW mit dem sehr hohen Wert von 5.03 auffällt. Dieser hohe Wert kommt wahrscheinlich aufgrund der Skala AWT1 zustande. Alle anderen Werte sind unauffällig.

Zum Abschluss der konfirmatorischen Faktorenanalyse werden die Diskrepanzen der empirisch geschätzten und der durch das Modell implizierten Korrelationen sowie Kovarianzen betrachtet (siehe Tabelle 3.13). Beim Vergleich der Korrelationen der empirischen Daten mit den durch das Modell geschätzten Daten beträgt die höchste Dis-

krepanz .20 zwischen der Skala ALV und EWT1. Dies entspricht einem Unterschied von 4 % Varianzaufklärung. Alle anderen Werte weisen nur geringe Diskrepanzen auf. Die höchste Diskrepanz bei den Kovarianzen zwischen den empirischen und den durch das Modell geschätzten Daten liefern die Skala ALV und AWT1. Auffällig ist, dass die Skala AWT1 bei den Kovarianzen die höchsten Diskrepanzen aufweist, dennoch sind diese als gering zu bezeichnen.

Tabelle 3.13

Diskrepanzen der empirisch geschätzten und der durch das Modell implizierten Korrelationen (unter der Diagonalen) und Kovarianzen

	LV	AWT1	AWT2	Einst.	ALV	EWT1	EWT2
LV		-2.89	0.70	-1.61	-1.43	0.04	2.90
AWT1	-.06		0.14	3.17	4.84	0.19	-3.35
AWT2	.02	.00		-0.36	-1.26	-1.12	0.64
Einst.	-.11	.13	-.02		0.02	0.88	-0.69
ALV	-.06	.11	-.04	.00		3.15	-0.48
EWT1	.00	.00	-.05	.10	.20		0.01
EWT2	.16	-.11	.03	-.08	-.03	.00	

Insgesamt sprechen die meisten Parameter für einen guten Modell-Fit, die Indizes hingegen nicht.

3.4.6 Das neue Leistungsprüfsystem (LPS-neu)

Für die Intelligenztestung wurde in dieser Arbeit das von Kreuzpointner (2010) überarbeitete Leistungsprüfsystem (LPS-neu) verwendet.

3.4.6.1 Beschreibung

Das LPS wurde 1962 von Horn veröffentlicht (1962/1983) und basiert – wie bereits erwähnt – auf dem Primärfaktorenmodell von Thurstone (1938). Entsprechend diesem Modell, in dem sich die allgemeine Intelligenz als eine Matrix aus verschiedenen voneinander unabhängigen Faktoren darstellen lässt, entwickelte Horn einen Test, der die von Thurstone genannten Primärfaktoren – verbal factor, reasoning, word fluency, perceptual speed, space, closure und accuracy – erfassen sollte. Das LPS ist für Kinder und Erwachsene im Alter von 9 bis 85 Jahren geeignet.

Bei dem in dieser Studie verwendeten Leistungstest handelt es sich um die revidierte Form des LPS von Kreuzpointner (2010). Durch die Revision wollte Kreuzpointner (2010) ein ökonomischeres Verfahren, unter Beibehaltung der Konzeption nach Horn, entwickeln und schlug dabei gleichzeitig das Drei-Ebenen-Modell von Carroll (1993) als theoretische Grundlage vor. Aus diesem Grund besteht das LPS-neu ebenfalls aus vier DIN A 4 Seiten, die auf einem DIN A3 Bogen gedruckt sind. Auch die Bearbeitung der einzelnen Items ist im LPS-neu direkt in der Vorlage möglich. Zugunsten der Übersichtlichkeit wurden sowohl die Instruktionen der einzelnen Subtests, anstatt vor jeder Aufgabe, auf der ersten Seite zusammengefasst als auch bei einigen Items Änderungen der Antwortmodalitäten vorgenommen. Die Versuchsleiterinstruktion wurde der Version von Horn ebenfalls entnommen und auf das LPS-neu angepasst. Nach Kreuzpointners (2010) Auswertungen diverser faktoranalytischer Studien, Interkorrelations- und Korrelationsberechnungen des LPS mit anderen Intelligenztests wurden die Subtests entsprechend inhaltlich verändert bzw. zusammengefasst oder weggelassen, sodass das LPS-neu aus nicht mehr fünfzehn, sondern elf Subtests besteht. Dadurch konnte die Bearbeitungsdauer auf 60 Minuten reduziert werden. Diese elf Subtests messen, wie eine konfirmatorische Faktorenanalyse von Kreuzpointner (2010) bestätigen konnte, entsprechend dem Drei-Ebenen-Modell von Carroll (1993), vier Faktoren kognitiver Fähigkeiten der zweiten Ebene – kristalline und fluide Intelligenz, visuelle Wahrnehmung und kognitive Geschwindigkeit – und einen Faktor, die allgemeine Intelligenz, g , auf oberster Ebene. Die Struktur des LPS-neu kann der Abbildung 3.3 entnommen werden.

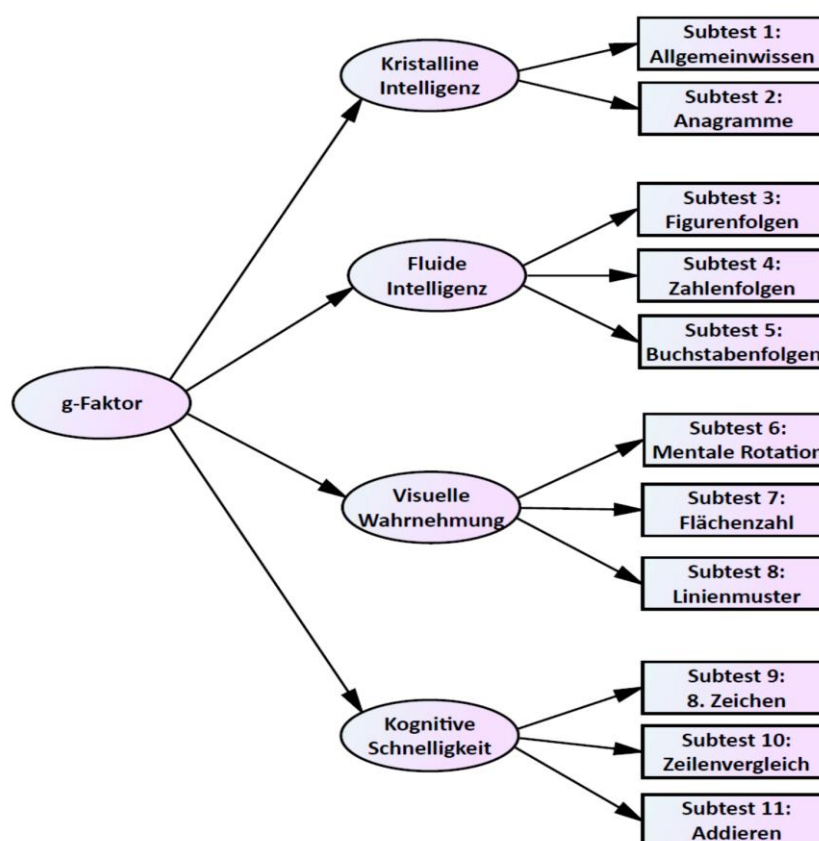


Abbildung 3.3

LPS-neu in der Struktur des Drei-Ebenen-Modells von Carroll (1993)

Für die Testungen an Schüler/-innen wurde die Instruktion des LPS-neu dahingehend verändert, dass anstatt der förmlichen Sie- die Du-Form verwendet wurde. Beide Versionen A und B des LPS-neu wurden in den Testungen verwendet. Diese sind in der Reihenfolge der Subtests identisch, fragen aber andere vergleichbare Items ab. Was sich hinter den einzelnen Untertests verbirgt, wird im Folgenden kurz skizziert.

Der erste Subtest besteht aus einer Liste von 60 Wörtern, die jeweils einen Druckfehler enthalten. Die Aufgabe der Probanden ist es nun, den falschen Buchstaben durch ein Kreuz zu kennzeichnen, wobei sie dafür insgesamt drei Minuten Zeit haben. Mit diesem Test wird die Allgemeinbildung getestet, da davon ausgegangen wird, dass Probanden mit einer hohen Allgemeinbildung mehr Wörter kennen und auch wissen, wie diese richtig geschrieben werden. Der zweite Subtest besteht aus 40 Wörtern, deren Buchstaben durcheinander angeordnet und zudem verstümmelt dargestellt sind. Das Ziel dieser Aufgabe besteht darin, den Anfangsbuchstaben des Wortes zu kennzeichnen. Dieser Untertest misst ebenfalls die Allgemeinbildung, aber über die sprachliche Kompetenz. Auch hier liegt das Zeitlimit bei drei Minuten. Subtest 1 entspricht dabei den Stratum I

Faktoren Verbal Ability sowie Lexical Knowledge von Carroll (1993). Für den Subtest 2 ist zusätzlich Carrolls (1993) Stratum I Faktor Reading Decoding erforderlich. Zusammen kann durch die ersten beiden Subtests die kristalline Intelligenz der Stratum II Ebene erfasst werden (siehe Abbildung 3.3).

Mithilfe der Subtests 3 bis 5 kann die fluide Intelligenz der Stratum II Ebene gemessen werden. Allen drei Tests ist gemeinsam, dass sie dem schlussfolgernden Denken zugeordnet werden können und aus jeweils 40 Items bestehen. Die Probanden haben dabei die Aufgabe, aus einer Reihe von acht Formen, neun Zahlen bzw. neun Buchstaben (Subtest 3, 4, 5) jeweils das Item anzukreuzen, welches nicht in die vorgegebene Systematik passt. Die ihnen dafür zur Verfügung stehende Zeit beträgt für den Subtest 3 drei Minuten und für die anderen beiden Tests je fünf Minuten.

Die Stratum II Ebene visuelle Wahrnehmung wird durch die Subtests 6 bis 8 gemessen. Dabei bestehen alle Tests aus jeweils 40 Items und nehmen je zwei (Subtest 6, 8) bzw. drei Minuten (Subtest 7) in Anspruch. In Untertest 6 müssen die Probanden von fünf in einer Zeile dargestellten Zeichen das spiegelverkehrte Zeichen erkennen. Der Subtest erfasst den Faktor Space und kann dem Stratum I Faktor räumliche Beziehung zugeordnet werden. Die Ermittlung der Flächenzahl eines dreidimensional präsentierten Körpers mit konstanten Flächenvorgaben von zwei bis zehn ist die Aufgabe in Subtest 7. Der Subtest kann also mit den Faktoren Visualisation und Imagination der Stratum I Ebene von Carroll (1993) gleichgesetzt werden. Im achten Subtest müssen die Probanden ein Linienmuster mit fünf dargestellten Mustern vergleichen und das Muster wählen, das sich im Linienmuster wiederfindet. Der Subtest kann Carolls Stratum I Faktoren Wahrnehmungsgeschwindigkeit und Räumliches Visualisieren zugeordnet werden.

Die Subtests 9 bis 11 gehören zur vierten kognitiven Kompetenz zweiter Ordnung, der kognitiven Schnelligkeit. Der Subtest 10, in welchem die Zeilen des Tests 9 mit denen von Test 10 verglichen werden und nicht übereinstimmende Zeichen in Test 10 markiert werden müssen, nimmt zwei Minuten in Anspruch. Von den 60 präsentierten Items sind 40 in mindestens einem Zeichen nicht identisch. Subtest 9, der erst im Anschluss bearbeitet wird, beinhaltet 40 Items und zur Bearbeitung haben die Probanden fünf Minuten Zeit. Bei diesem Test sollen die Probanden die Spalte 9 von oben nach unten durchgehen und zunächst jede achte „0“ markieren. Am Ende angelangt, beginnen sie wieder von Neuem, und zwar zählen und markieren sie dann jede achte „1“, „2“, „3“ usw. Die beiden Subtests 9 und 10 erheben dabei das Wahrnehmungstempo, Subtest 9 zusätzlich auch die Sorgfalt bei der Bearbeitung. Der letzte Subtest 11, der die Kon-

zentrationenleistungsfähigkeit erfasst, besteht aus insgesamt 80 Items und die Probanden haben hierfür insgesamt sechs Minuten Zeit. Aufgabe ist es, die zehn präsentierten Zahlen zu addieren und die Einerziffer des Ergebnisses anzukreuzen.

3.4.6.2 Testanalyse

In dieser Studie wurde sowohl die Version A als auch B des LPS-neu verwendet, weshalb zunächst eine Analyse beider Versionen getrennt voneinander vorgenommen wurde. Zuerst wurde eine Reliabilitätsanalyse nach Cronbach's α und der Split-Half-Koeffizienten für die einzelnen Subtests der beiden Versionen durchgeführt (siehe Tabelle 3.14). Die Ergebnisse zu Cronbach's α geben an, wie stark die Items eines jeden Subtests das Gleiche messen. Diese internen Konsistenzen liegen dabei für die Subtests der Version A zwischen .84 und .96 sowie für die Version B zwischen .70 und .96 auf einem hohen Niveau. Zur weiteren Prüfung der Reliabilität sind in Tabelle 3.14 die Split-Half-Koeffizienten aufgeführt. Die nach Spearman-Brown-Formel korrigierten Halbierungsreliabilitäten sind für die Subtests der Version A zwischen .90 und .98 sehr zufriedenstellend. Die nach der Spearman-Brown-Formel korrigierten Halbierungsreliabilitäten fallen für die Subtests der Version B zwischen .84 und .99 etwas niedriger aus als für die Version A.

Aus der Tabelle 3.14 können die Mittelwerte und die Standardabweichungen der LPS-neu Subtests und des Gesamtwertes sowie die Schiefe und Kurtosis (Exzess) der Verteilungen für die Gesamtstichprobe, getrennt nach Version A und B, entnommen werden. Die Kurtosis liegt bei den meisten Subtests annähernd bei 0. Nur bei den Subtests 5, 7 und 10 der Version A und bei den Subtests 3, 4, 5 und 7 der Version B liegt eine etwas erhöhte Kurtosis und damit eine steilgipfligere Verteilung vor. Die meisten Subtests beider Versionen sowie der Gesamtscore der Version B weisen eine geringe Schiefe von $< .30$ auf. Lediglich bei Subtest 7 beider Versionen und bei Subtest 11 der Version B fällt eine hohe Schiefe auf. Da jedoch die Schiefe und Kurtosis der Subtests von der Normalverteilung nur geringfügig abweichen, kann davon ausgegangen werden, dass die erzielten Punktwerte im LPS-neu annähernd normalverteilt sind. Dies unterstreichen auch die Abbildungen im Anhang C (siehe Abbildung C-1; C-2), denen die Verteilungen der LPS-neu Subtests und des LPS-Gesamtwertes für die Version A und B entnommen werden können.

Tabelle 3.14

Deskriptive Statistik der Versionen A und B des LPS-neu, Reliabilitätsanalyse nach Cronbach's Alpha und Split-Half-Koeffizient, Schiefe und Kurtosis

Version A	<i>M</i>	<i>SD</i>	α^I	r_{tt}^I	Schiefe	Kurtosis
Subtest 1	25.60	6.92	.91	.94	-.13	-.09
Subtest 2	19.00	6.95	.90	.92	-.02	-.59
Subtest 3	20.60	3.30	.84	.90	-.10	.20
Subtest 4	19.59	3.68	.87	.93	-.19	-.13
Subtest 5	21.78	3.13	.87	.90	-.75	2.82
Subtest 6	19.67	8.19	.91	.97	.05	-.00
Subtest 7	26.24	5.34	.89	.97	-1.06	2.58
Subtest 8	23.98	5.89	.90	.93	-.33	.39
Subtest 9	9.34	4.23	.80	.92	.13	-.69
Subtest 10	31.05	8.79	.95	.99	.43	2.09
Subtest 11	11.37	5.67	.96	.98	-.06	-.52
Gesamtwert	228.10	34.25	.86	.92	-.47	.41
Version B	<i>M</i>	<i>SD</i>	α^I	r_{tt}^I	Schiefe	Kurtosis
Subtest 1	27.80	7.21	.89	.90	.24	-.27
Subtest 2	19.32	7.97	.91	.95	.31	-.44
Subtest 3	23.19	4.49	.85	.93	-.48	1.54
Subtest 4	16.77	4.19	.89	.94	.24	1.25
Subtest 5	20.36	3.87	.87	.91	-.45	1.40
Subtest 6	21.50	7.30	.91	.96	.17	.26
Subtest 7	27.91	5.44	.90	.98	-.98	1.47
Subtest 8	23.30	6.86	.91	.96	.14	.18
Subtest 9	8.83	4.10	.79	.94	.11	-.26
Subtest 10	28.75	10.21	.96	.99	-.02	.97
Subtest 11	8.95	3.78	.70	.84	-.96	.05
Gesamtwert	226.49	36.19	.83	.92	.02	.17

Anmerkungen. *M* = Mittelwert; *SD* = Standardabweichung; α = Cronbach's Alpha; r_{tt} = korrigierter Split-Half-Koeffizient nach Spearman-Brown-Formel; ^ITeilung in zwei gleichgroße Subtests (ungerade vs. gerade); beim Gesamtwert wurde jeweils die Profilreliabilität: $\text{prof } r_{tt} = \frac{\bar{r}_{tt} - \bar{r}_{st}}{1 - \bar{r}_{st}}$ berechnet (Lienert & Raatz, 1998, S. 324)

Die Subtests des LPS-neu korrelieren in geringem bis mittlerem positiven Ausmaß zwischen $r = .44$ (Subtest 9, 8. Zeichen) und $r = .65$ (Subtest 1, Allgemeinwissen) bei Version A und zwischen $r = .37$ (Subtest 11, Addieren) und $r = .62$ (Subtest 6 und 8, Mentale Rotation und Linienmuster) bei Version B mit dem LPS-neu Gesamtergebnis (siehe Tabelle 3.15).

Tabelle 3.15

Zusammenhänge zwischen den Subtests des LPS-neu (Version A und B)

Subtests Version A	1	2	3	4	5	6	7	8	9	10	11
1	-	.49**	.30**	.23**	.46**	.29**	.16*	.22**	.18*	.19**	.28**
2		-	.35**	.28**	.36**	.17*	.18*	.20**	.21**	.26**	.28**
3			-	.31**	.43**	.20**	.38**	.30**	.20**	.23**	.20**
4				-	.46**	.08	.24**	.23**	.19*	.05	.31**
5					-	.28**	.30**	.33**	.18*	.20**	.27**
6						-	.21**	.24**	.07	.09	.07
7							-	.37**	.27**	.17*	.20**
8								-	.11	.21**	.17**
9									-	.27**	.22**
10										-	.15**
11											-
Gesamtwert	.65**	.64**	.58**	.47**	.64**	.50**	.54**	.55**	.44**	.54**	.50**
Subtests Version B	1	2	3	4	5	6	7	8	9	10	11
1	-	.43**	.16*	.20**	.22**	.31**	.17*	.28**	.03	.18*	.14
2		-	.22**	.26**	.36**	.23**	.18*	.31**	.15*	.06	.11
3			-	.27**	.38**	.30**	.36**	.33**	.10	.27**	.13
4				-	.41**	.22**	.22**	.29**	.20**	.17*	.31**
5					-	.25**	.22**	.30**	.20**	.22**	.22**
6						-	.18*	.37**	.17*	.32**	.15*
7							-	.41**	.22**	.18*	.21**
8								-	.11	.16*	.10
9									-	.25**	.24**
10										-	.15*
11											-
Gesamtwert	.57**	.58**	.55**	.52**	.57**	.62**	.53**	.62**	.39**	.58**	.37**

Anmerkungen: **) Die Korrelation ist auf dem .01 Niveau signifikant; *) Die Korrelation ist auf dem .05 Niveau signifikant

Die höchsten Zusammenhänge zwischen den Subtests weisen bei beiden LPS-neu Versionen die Zahlen- und Buchstabenfolgen (Subtests 4 und 5), das Allgemeinwissen und die Anagramme (Subtests 1 und 2) sowie die Flächenzahl und die Linienmuster (Subtests 7 und 8) auf (siehe Tabelle 3.15). Zudem korreliert bei Version A der Subtest 5 hoch mit Subtest 1 (Allgemeinwissen) und Subtest 3 (Figurenfolgen).

Um zu überprüfen, ob die beiden Versionen des LPS-neu äquivalent sind und die weiteren Berechnungen gemeinsam ausgewertet werden können, wurden bei beiden Versionen für die Subtests, für den Gesamtscore und für die vier Dimensionen des LPS-neu – kristalline Intelligenz (LPS_Gc), fluide Intelligenz (LPS_Gf), visuelle Wahrnehmung (LPS_Gv), kognitive Schnelligkeit (LPS_Gs) – Mittelwertsvergleiche berechnet. Der Tabelle 3.16 können die Berechnungen entnommen werden. Bei Mittelwertsvergleichen, bei denen der Signifikanztests – die Prüfung auf Varianzhomogenität – signifikant war, wurde der Welch-Test für die Analyse angewandt. Alle anderen wurden mithilfe des *t*-Tests berechnet.

Tabelle 3.16

Mittelwerts- und Varianzvergleich der Intelligenztestergebnisse, der einzelnen Subtests und der vier Dimensionen des LPS-neu zwischen Version A und B, mithilfe des Levene-Tests

	<i>F</i> - Wert	<i>p</i>	<i>t</i> - Wert	<i>df</i>	<i>p</i>	<i>g</i>
Subtest 1	.77	.38	-2.97	359	.00	-.31
Subtest 2	3.02	.08	-.41	354.83	.68	-.04
Subtest 3	10.02	.00	-6.26	334.09	.00	-.66
Subtest 4	.95	.33	6.78	359	.00	.71
Subtest 5	11.52	.00	3.84	347.37	.00	.40
Subtest 6	1.97	.16	-2.37	359	.02	-.24
Subtest 7	.34	.56	-2.96	359	.00	-.31
Subtest 8	2.65	.10	1.01	359	.31	.15
Subtest 9	1.37	.24	1.27	359	.20	.12
Subtest 10	4.56	.03	2.42	352.15	.02	.24
Subtest 11	44.82	.00	4.91	286.90	.00	.50
Gesamtwert	.27	.60	.44	359	.66	.05
LPS_Gc	.47	.49	-1.93	359	.05	-.20
LPS_Gf	5.31	.02	1.81	349.86	.07	.19
LPS_Gv	1.05	.31	-1.95	359	.05	-.21
LPS_Gs	.01	.91	3.91	359	.00	.41

Anmerkungen. *F* = Testwert; *p* = Signifikanzwert; *t* = Testwert; *g* = Hedge *g* (Effektstärke)

Dabei ergab der Mittelwertsvergleich des Gesamtscores der Versionen A und B mithilfe des *t*-Testes bei unabhängigen Stichproben ($t(359) = .44$, $p = .66$, $g = .05$), dass sich die Mittelwerte nicht signifikant voneinander unterscheiden (siehe Tabelle 3.16). Aus diesem Grund wurden die beiden Versionen, zumindest den Gesamtscore betreffend, zusammen ausgewertet. Wie allerdings der Tabelle 3.16 zu entnehmen ist, unterscheiden

sich zum einen einige Subtests (1, 3 - 7, 10, 11) der beiden Versionen signifikant voneinander und zum anderen ist auch die Dimension kognitive Schnelligkeit des LPS-neu signifikant unterschiedlich. Alle weiteren Berechnungen, bei denen die Subtests bzw. die vier Dimensionen des LPS-neu Relevanz fanden, wurden aufgrund dessen für beide Versionen getrennt berechnet.

3.5 Prüfung der Unterschiedlichkeit zwischen verschiedenen Gruppen

Um die Vergleichbarkeit einzelner Gruppen bzw. des Geschlechts und des Alters zu gewährleisten, wurde überprüft, ob sich diese in den einzelnen Leistungen bzw. in der Höhe ihrer Motivation unterscheiden.

Aus diesem Grund wurde zunächst geprüft, ob das Geschlecht bzw. das Alter einen Einfluss auf die Höhe der Testmotivation, die Intelligenz bzw. die Leistungsmotivation haben.

Dabei erbrachte eine Analyse der Daten für die Testmotivation – bei der 332 Versuchspersonen berücksichtigt wurden (siehe Kapitel 3.3) – mithilfe von univariaten Varianzanalysen zur Untersuchung von Haupt- und Interaktionseffekten hinsichtlich des Alters und des Geschlechts mit der Testmotivation folgende Ergebnisse: Es ergab sich weder ein signifikanter Haupteffekt des Geschlechts, noch konnte dieser für das Alter ausfindig gemacht werden (siehe Tabelle 3.17).

Tabelle 3.17

Univariate Varianzanalyse der Testmotivation mit Alter und Geschlecht

TM	<i>F</i> (TMg)	<i>p</i> (TMg)	η^2 (TMg)	<i>F</i> (TMv)	<i>p</i> (TMv)	η^2 (TMv)	<i>F</i> (TMn)	<i>p</i> (TMn)	η^2 (TMn)
Alter	1.87	.12	.02	1.42	.23	.02	1.68	.16	.02
Geschl.	1.27	.26	.00	.55	.46	.00	1.71	.19	.01
Alter * Geschl.	1.06	.37	.01	.55	.70	.01	1.52	.20	.02

Anmerkungen: *F* = Testwert; *p* = Signifikanzwert; η^2 = Effektgröße partielles Eta-Quadrat;

Dies unterstreicht, dass sich weder die Geschlechter hinsichtlich ihrer Testmotivation unterscheiden noch, dass das Alter darauf einen Einfluss hat. Darüber hinaus ergab eine Analyse der Testmotivation, getrennt für den Fragebogen vor und nach dem Test – berücksichtigt wurden nur die Skalen, die jeweils vor bzw. nach dem Test erfasst wurden –, dass auch hier weder ein signifikanter Haupteffekt des Geschlechts noch einer für das

Alter vorliegt (siehe Tabelle 3.17). Der Mittelwert der gesamten Testmotivation liegt bei den Mädchen bei 166.75 ($SD = 25.14$) und bei den Jungen bei 166.23 ($SD = 25.99$). Wie aus der Abbildung 3.4 deutlich wird, weist die Gruppe der 18-jährigen Mädchen den niedrigsten Mittelwert von 154.21 ($SD = 19.67$) der gesamten Testmotivation auf. Die höchste gesamte Testmotivation zeigt sich bei den 17-jährigen Mädchen und bei gleichaltrigen Jungen. Die Mittelwerte für die Testmotivation vor und nach dem Test der Mädchen ($M = 97.96$; $SD = 15.12$ vs. $M = 68.78$; $SD = 12.78$) sind mit denen der Jungen fast identisch ($M = 97.46$; $SD = 15.37$ vs. $M = 68.77$; $SD = 13.79$).

Der Abbildung 3.4 kann ebenfalls entnommen werden, dass sich Mädchen und Jungen zwischen 14 und 16 hinsichtlich ihrer Testmotivation kaum unterscheiden sowohl vorher als auch nachher oder im Gesamten. Die 17-Jährigen weisen bei den Schülerinnen und den Schülern, im Vergleich zu den jüngeren Altersklassen, eine höhere Testmotivation auf. Bei den 18-Jährigen sinkt die Testmotivation vor dem Test und insgesamt hingegen wieder. Nur bei der Testmotivation nach dem Test weisen die 18-jährigen Schüler eine höhere Testmotivation auf.

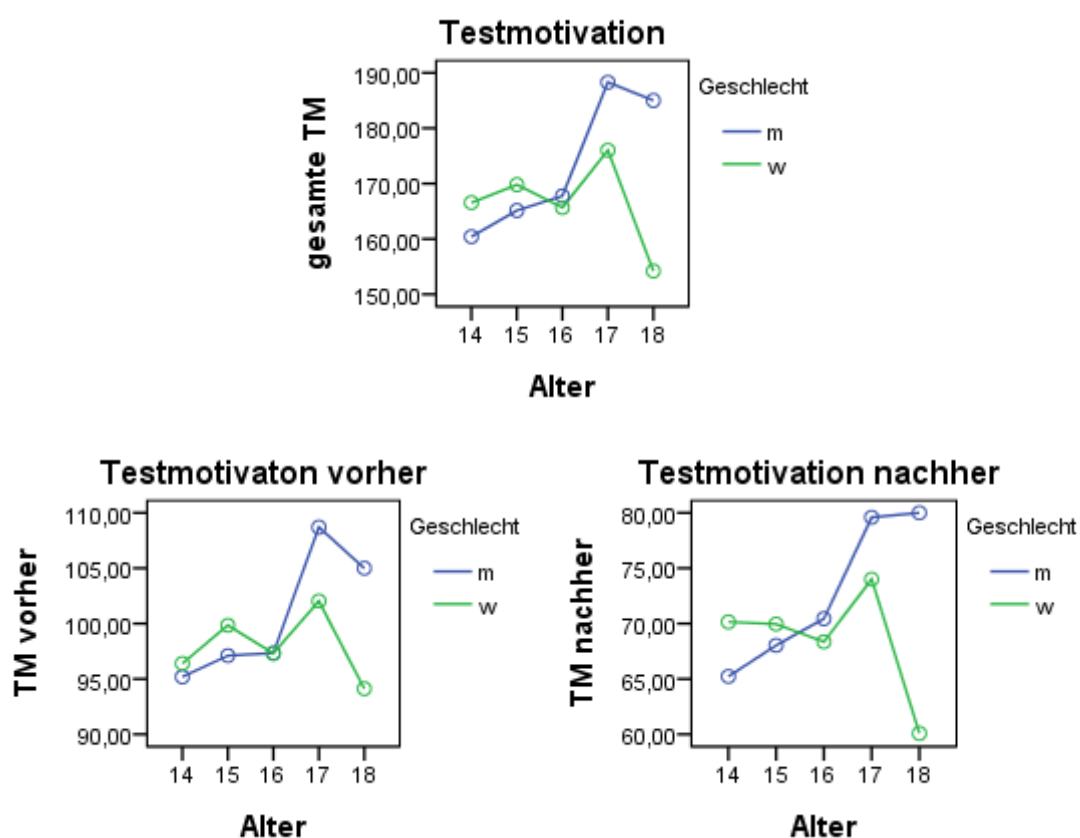


Abbildung 3.4

Univariate Varianzanalyse der TMg bzw. TMv sowie TMn mit Alter und Geschlecht

Da die Testmotivation vor und nach dem Test erhoben wurde, war es empfehlenswert zu überprüfen, ob diese, infolge der Testung, signifikante Unterschiede aufweisen. Um einen Mittelwertsvergleich zwischen der Testmotivation vor und nach dem Test zu ermöglichen, mussten die Daten zunächst bearbeitet werden, da sie sich aus einer unterschiedlichen hohen Anzahl von Items zusammensetzen. Die Testwerte der Motivation, vorher und nachher, aller Versuchspersonen wurde dafür durch die Anzahl der jeweiligen Items geteilt. So wurden die Werte der Testmotivation, die vor dem Test erhoben wurden, durch 31 und die Werte der Motivation nach dem Test durch 22 dividiert. Die schließlich erhaltenen Werte wurden mit einem t -Test bei gepaarten Stichproben geprüft. Dieser ergab, dass sich die Testmotivation vor dem Test nicht signifikant von der Testmotivation nach dem Test unterscheidet ($t(331) = .89$, $p = .38$, $g = .04$). Dementsprechend kann davon ausgegangen werden, dass die Versuchspersonen vor dem Test ($M = 3.15$; $SD = .49$) genauso motiviert waren wie nach dem Test ($M = 3.13$; $SD = .61$), was überdies durch die signifikante Korrelation ($r = .60$; $p = .00$) Bestätigung findet.

Auch die allgemeine Intelligenzleistung wurde mithilfe einer univariaten Varianzanalyse zur Untersuchung von Haupt- und Interaktionseffekten hinsichtlich des Alters und des Geschlechts untersucht. Dabei zeigte sich weder ein signifikanter Haupteffekt des Geschlechts noch konnte dieser für das Alter ausfindig gemacht werden (siehe Tabelle 3.18). Somit kann festgehalten werden, dass weder das Geschlecht noch das Alter einen Einfluss auf die Intelligenzleistung hat.

Tabelle 3.18

Univariate Varianzanalyse des LPS-neu mit Alter und Geschlecht

LPS_G	F	p	$p\eta^2$
Alter	1.30	.27	.02
Geschlecht	.01	.92	.00
Alter * Geschlecht	.23	.87	.00

Anmerkungen. F = Testwert; p = Signifikanzwert; $p\eta^2$ = Effektgröße partielles Eta-Quadrat

Die Mittelwerte der Mädchen liegen bei 226.55 ($SD = 35.50$) und bei den Jungen bei 227.83 ($SD = 35.07$). Abbildung 3.5 zeigt, dass die 16-jährigen Schüler mit 232.43 ($SD = 37.59$) den höchsten Mittelwert in der allgemeinen Intelligenzleistung aufweisen. Den niedrigsten Mittelwert hat der 19-jährige Schüler mit 198.00, gefolgt von den 18-jährigen Mädchen mit 210.54 ($SD = 37.33$). Dabei unterscheiden sich die Leistungen von Mädchen und Jungen sowie abhängig vom Alter nicht signifikant.

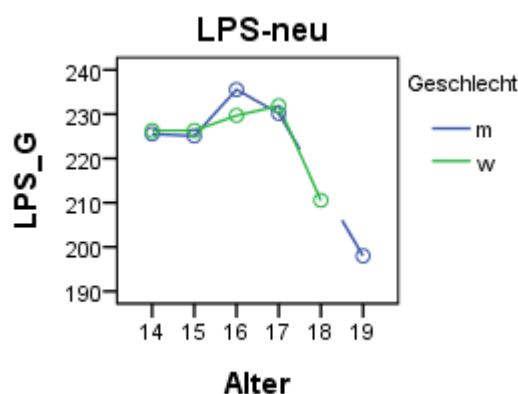


Abbildung 3.5

Univariate Varianzanalyse des LPS-Gesamtwertes mit Alter und Geschlecht

Das genaue Alter in Monaten wurde ebenfalls der Analyse unterzogen, und zwar bezüglich der Untersuchung von Haupt- und Interaktionseffekten der Intelligenzleistung zusammen mit dem Geschlecht. Auch hier ergaben sich keine signifikanten Ergebnisse ($F(34) = .72$; $p = .88$; $\eta^2 = .08$).

Zudem wurde überprüft, ob sich die Geschlechter hinsichtlich der einzelnen Subtests des LPS-neu voneinander unterscheiden. Die Ergebnisse, getrennt für die beiden Versionen A und B, können den Tabellen 3.19 und 3.20 entnommen werden.

Tabelle 3.19

Mittelwertsvergleich der Intelligenztestergebnisse der einzelnen Subtests des LPS-neu zwischen Schülern und Schülerinnen (Version A)

Version A	Schüler (n = 100)	Schülerinnen (n = 78)				
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>t</i> -Wert	<i>df</i>	<i>p</i>	<i>g</i>
Subtest 1	25.84 (7.32)	25.28 (6.40)	.53	176	.60	.08
Subtest 2	18.84 (6.76)	19.21 (7.23)	-.35	176	.73	-.05
Subtest 3	20.72 (3.74)	20.45 (3.21)	.54	176	.59	.08
Subtest 4	20.10 (3.89)	18.94 (3.30)	2.12	176	.04	.32
Subtest 5	21.84 (3.28)	21.69 (2.93)	.31	176	.76	.05
Subtest 6	20.01 (7.75)	18.97 (8.97)	.83	176	.41	.14
Subtest 7	26.32 (5.12)	26.13 (5.65)	.24	176	.81	.04
Subtest 8	24.46 (6.09)	23.37 (5.60)	1.22	176	.22	.19
Subtest 9	8.75 (4.17)	10.10 (4.22)	-2.14	176	.03	-.32
Subtest 10	29.10 (7.28)	33.55 (9.91)	-3.45	176	.00	-.52
Subtest 11	12.02 (5.89)	10.54 (5.30)	1.74	176	.08	.26

Anmerkungen. *M* = Mittelwert; *SD* = Standardabweichung; *t* = Testwert; *df* = Freiheitsgrade; *p* = Signifikanzwert; *g* = Hedge *g* (Effektgröße)

Tabelle 3.20

Mittelwertsvergleich der Intelligenztestergebnisse der einzelnen Subtests des LPS-neu zwischen Schülern und Schülerinnen (Version B)

Version B	Schüler (n= 107)	Schülerinnen (n= 76)				
	<i>M (SD)</i>	<i>M (SD)</i>	<i>t-Wert</i>	<i>df</i>	<i>p</i>	<i>g</i>
Subtest 1	29.43 (7.35)	25.51 (6.37)	3.84	173.86	.00	.56
Subtest 2	19.66 (7.83)	18.84 (8.20)	.69	181	.49	.10
Subtest 3	23.45 (4.51)	22.83 (4.47)	.92	181	.36	.14
Subtest 4	17.10 (4.34)	16.30 (3.96)	1.27	181	.20	.19
Subtest 5	20.09 (3.94)	20.72 (3.77)	-1.09	181	.28	-.16
Subtest 6	21.18 (6.97)	21.96 (7.77)	-.71	181	.48	-.11
Subtest 7	28.64 (5.10)	26.89 (5.76)	2.16	181	.03	.33
Subtest 8	23.29 (6.48)	23.32 (7.42)	-.03	181	.98	-.00
Subtest 9	8.32 (3.77)	9.43 (4.54)	-1.81	181	.07	-.27
Subtest 10	27.36 (9.06)	30.33 (11.89)	-1.91	181	.06	-.29
Subtest 11	9.14 (3.41)	8.68 (3.34)	.90	181	.34	.14

Anmerkungen. *M* = Mittelwert; *SD* = Standardabweichung; *t* = Testwert; *df* = Freiheitsgrade; *p* = Signifikanzwert; *g* = Hedge *g* (Effektgröße)

Es wird ersichtlich, dass außer in Subtest 1 der Version B des LPS-neu der *t*-Test verwendet wurde. In diesem Subtest wurde der Signifikanztest – Prüfung auf Varianzhomogenität – signifikant (Subtest 1: $F = 4.15$; $p = .04$), weshalb hier der Welch-Test für die Analyse angewandt wurde. Von den elf Subtests der Version A (siehe Tabelle 3.19) zeigen drei signifikante Mittelwertsunterschiede zwischen Schülern und Schülerinnen. Die Schüler erzielen im Subtest 4 (Mittlere Differenz = 1.16; *SD* der Differenz = .55) bessere Werte als die Schülerinnen. Die Schülerinnen schneiden hingegen in den Subtests 9 (Mittlere Differenz = 1.35; *SD* der Differenz = .63) und 10 (Mittlere Differenz = 4.45; *SD* der Differenz = 1.29) mit besseren Ergebnissen ab. Die Version B des LPS-neu weist zwei signifikante Mittelwertsunterschiede zwischen Schülern und Schülerinnen auf (siehe Tabelle 3.20): So erzielen die Schüler sowohl im Subtest 1 (Mittlere Differenz = 3.92; *SD* der Differenz = 1.05) als auch im Subtest 7 (Mittlere Differenz = 1.74; *SD* der Differenz = .81) bessere Ergebnisse als die Schülerinnen.

Da auch die Daten der Klasse des vorherigen Schuljahres erhoben wurden, ist interessant, ob sich diejenigen, die eine Klasse wiederholen mussten, in ihren Intelligenzleistungen von den Nicht-Wiederholern unterscheiden. Für diese Berechnungen blieb die Grund- und Mittelschule unberücksichtigt, da die Daten der Klasse des vorherigen

rigen Schuljahres an dieser Schule nicht erhoben wurden. Auch die drei Versuchspersonen, die zu diesem Sachverhalt keine Angaben machten, konnten in den Berechnungen nicht berücksichtigt werden. Verglichen wurden in der verbleibenden Stichprobe ($n = 320$) die Wiederholer mit den Nicht-Wiederholern. Es ergibt sich für die Wiederholer ein LPS-G Mittelwert von 217.61 ($SD = 40.54$) und für die Nicht-Wiederholer ein Wert von 232.12 ($SD = 32.29$). Mithilfe eines t -Testes wurde überprüft, ob sich die beiden Gruppen in ihren Intelligenzleistungen signifikant voneinander unterscheiden. Dabei wurde der Levene-Test auf Varianzhomogenität ($F = 1.28$; $p = .27$) nicht signifikant. Für den t -Test ergab sich ein signifikanter Unterschied zwischen den Gruppen ($t(318) = 2.32$, $p = .02$; $g = .44$). Es kann also davon ausgegangen werden, dass sich die beiden Gruppen in ihren Intelligenzleistungen signifikant unterscheiden, wobei die Gruppe der Nicht-Wiederholer bessere Ergebnisse im LPS-neu erzielt.

Auch die vier standardisierten Leistungsmotive – HH, HL, LH und LL (errechnet aus den beiden Fragebögen zur Leistungsmotivation) – wurden mithilfe einer univariaten Varianzanalyse zur Untersuchung von Haupt- und Interaktionseffekten hinsichtlich des Alters und des Geschlechts untersucht. Dabei ergab sich weder ein signifikanter Haupteffekt des Geschlechts noch konnte dieser für das Alter ausfindig gemacht werden (siehe Tabelle 3.21). Es unterscheiden sich somit die Geschlechter nicht hinsichtlich ihrer Leistungsmotivausprägung und auch das Alter übt keinen Einfluss auf die Leistungsmotive aus.

Tabelle 3.21

Univariate Varianzanalyse der Leistungsmotive mit Alter und Geschlecht ($n = 313$)

LM	F (HH)	p (HH)	η^2 (HH)	F (HL)	p (HL)	η^2 (HL)	F (LH)	p (LH)	η^2 (LH)	F (LL)	p (LL)	η^2 (LL)
Alter	.31	.87	.00	1.80	.13	.02	1.75	.14	.02	1.25	.29	.02
Gesch.	.44	.51	.00	1.18	.28	.00	1.06	.31	.00	.01	.92	.00
Alt.*Gesch.	1.39	.24	.02	.60	.66	.01	.18	.95	.00	1.00	.41	.01

Anmerkungen: F = Testwert; p = Signifikanzwert; η^2 = Effektgröße partielles Eta-Quadrat;

Dass sich für die Leistungsmotivtypen kein signifikanter Haupteffekt des Geschlechts und des Alters finden lässt, kann auch der Abbildung 3.6 entnommen werden. So weisen zwar beispielsweise die 17-jährigen Schüler bei den Motiven HH und LH die höchsten Werte auf, diese unterscheiden sich jedoch nicht signifikant von den Werten der anderen Altersklassen bzw. von denen der Schülerinnen.

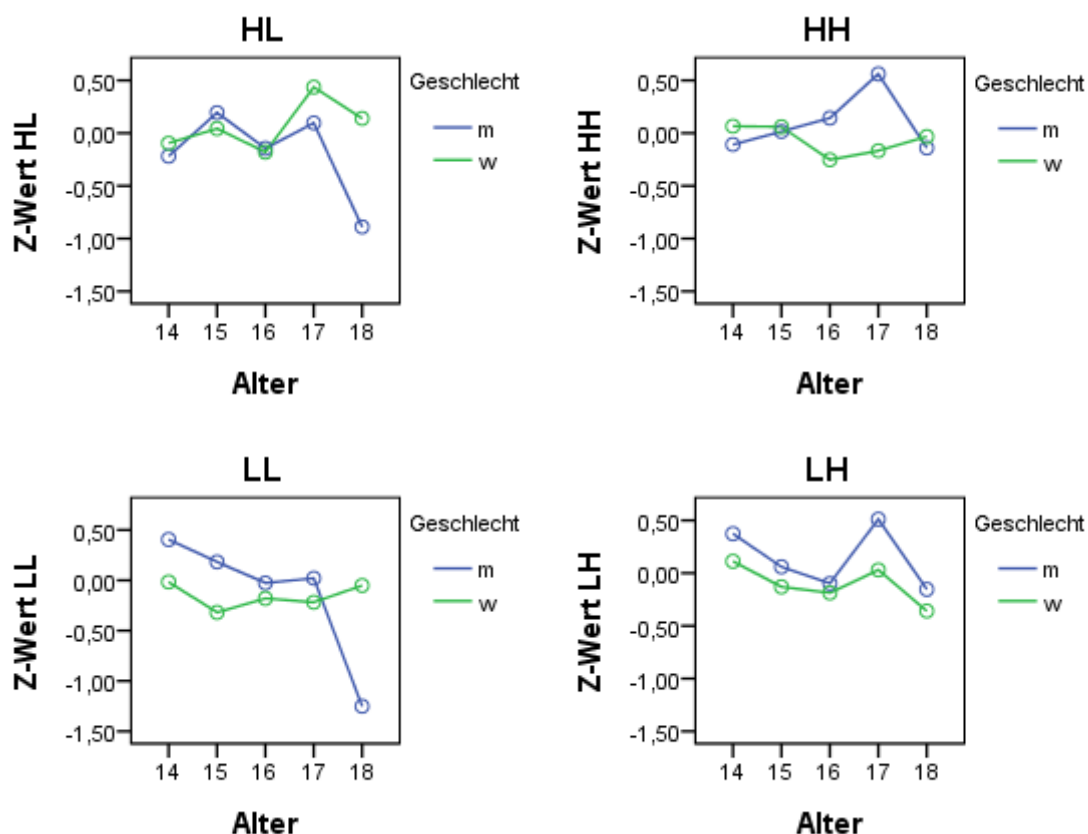


Abbildung 3.6

Univariate Varianzanalyse Leistungsmotive mit Alter und Geschlecht

Zusammenfassend lässt sich festhalten, dass weder Alters- noch Geschlechtseffekte bei der allgemeinen Intelligenz, bei den Leistungsmotiven und bei der Testmotivation – gesamt, vorher und nachher – gefunden werden konnten. Lediglich bei einzelnen Subtests des LPS-neu unterscheiden sich die Geschlechter signifikant in ihren Leistungen. Diejenigen Schüler/-innen, die eine Klasse im vorherigen Schuljahr wiederholen mussten, liefern ebenfalls signifikant unterschiedliche Leistungen im LPS-neu im Vergleich zu den Nicht-Wiederholern. So erzielen sie schlechtere Ergebnisse in der allgemeinen Intelligenz, als Schüler/-innen, die keine Klasse wiederholten.

4. Ergebnisse

Im Folgenden werden die Ergebnisse für die im Kapitel 2 dargestellten Hypothesen und Fragestellungen vorgestellt. Zu Beginn werden die Ergebnisse der einzelnen Konstrukte und im Anschluss daran die Zusammenhänge zwischen diesen dargestellt.

4.1 Leistungsmotiv

- H1: Die Einbindung der Leistungsmotivkomponenten des RLMI-K/J und des AMG-S K-J in das quadripolare Modell der Leistungsmotivation von Covington et al. (Covington & Omelich, 1991; Covington & Roberts, 1994) ist erfolgreich.*
- H2: Die Leistungsmotive HE und FM fungieren als unabhängige Konstrukte.*

Voraussetzung für die Erfüllung der oben genannten Hypothese 1 ist, dass die Leistungsmotive HM und FE des RLMI-K/J miteinander korrelieren und sich zu der Motivkomponente LL zusammenfügen lassen. Zudem sollte das Leistungsmotiv HE des RLMI-K/J mit HE des AMG-S K-J korrelieren, sodass diese sich zu HL zusammenfassen lassen. Auch die Komponenten FM des RLMI-KJ und FMp des AMG-S K-J sollten korrelieren und sich zu LH zusammenfassen lassen. Ferner sollte die FMa des AMG-S K-J der Komponente HH entsprechen. Zunächst mussten jedoch die Daten bereinigt werden. Versuchspersonen, bei denen mehr als ein Item zur Vervollständigung einer Skala fehlte, wurden in den Berechnungen nicht berücksichtigt, sodass für die weiteren Rechnungen 313 Versuchspersonen einbezogen werden konnten. Insgesamt wurden 24 Items durch Mittelwerte ersetzt.

Wie der Tabelle 4.1 zu entnehmen ist, korrelieren die Motive HM und FE mit $r = .79$ ($p = .00$) signifikant und lassen sich somit zu LL zusammenfassen. Die Motive HE des AMG-S K-J und HE des RLMI-K/J korrelieren signifikant mit $r = .28$ ($p = .00$) und können sich somit ebenfalls zu einer Komponente HL verbinden lassen. Die FMp des AMG-S K-J korreliert mit der FM des RLMI-K/J jedoch nicht signifikant. Die FM des RLMI-K/J weist hingegen signifikante Werte mit den Komponenten HM und FE auf, was einer Einbindung zu einem gemeinsamen Konstrukt HL entgegenspricht.

Tabelle 4.1

Korrelationen der einzelnen Leistungsmotive des AMG-S K-J und des RLMI-K/J

	HE AMG-S	FMa AMG-S	FMp AMG-S	HE RLMI	FM RLMI	FE RLMI	HM RLMI
HE AMG-S	-	.44**	-.00	.28**	.01	.04	-.02
FMa AMG-S		-	.22**	.12*	.19**	.03	.04
FMp AMG-S			-	-.20**	.11	.12*	.12*
HE RLMI				-	-.11*	-.04	-.13**
FM RLMI					-	.46**	.48**
FE RLMI						-	.79**
HM RLMI							-

Anmerkungen. **) Die Korrelation ist auf dem .01 Niveau signifikant; *) Die Korrelation ist auf dem .05 Niveau signifikant; auf die Bezeichnung: K-J bzw. K/J wurde aufgrund der Übersichtlichkeit in der Tabelle verzichtet

Darüber hinaus korreliert die Komponente FMa des AMG-S K-J sowohl mit HE desselben Messinstruments als auch mit der FM des RLMI-K/J und mit der FMp des AMG-S K-J signifikant (siehe Tabelle 4.1). Ob diese sich zu einer eigenständigen Komponente einbinden lässt und ob die Einbindung der Komponenten entsprechend des Modelles möglich ist, werden die anschließend dargestellte Faktorenanalyse, die multidimensionale Skalierung und die Clusteranalyse zeigen.

Zunächst wurde eine Faktorenanalyse für alle sieben Komponenten, die die Leistungsmotive erfassen, berechnet. Diese lassen sich durch drei Faktoren mit einer erklärten Gesamtvarianz von 71.86 % darstellen. Die rotierte Komponentenmatrix ist in der Tabelle 4.2 dargestellt.

Tabelle 4.2

Komponentenmatrix der Leistungsmotive des AMG-S K-J und des RLMI-K/J

	1	2	3
HE RLMI/KJ	-.09	.42	-.69
FM RLMI/KJ	.73	.11	.14
FE RLMI/KJ	.91	.00	-.01
HM RLMI/KJ	.91	-.06	.06
HE AMG-S K-J	.01	.81	-.20
FMp AMG-S K-J	.07	.23	.81
FMa AMG-S K-J	.06	.83	.24

Aus der Tabelle 4.2 geht hervor, dass die Komponenten HM, FE und FM des RLMI-K/J durch den gleichen Faktor, hier Faktor 1, erklärt werden können, was sie entsprechend dem Modell nicht sollten. Die beiden Motivkomponenten HE sowohl des RLMI-K/J als auch des AMG-S K-J lassen sich am besten durch den zweiten Faktor erklären. Die Komponente FMp wird durch den dritten Faktor erklärt, der bei den beiden HE-Komponenten negativ und somit in die andere Richtung gepoolt ist. Auch die Motivkomponente FMa des AMG-S K-J weist einen sehr hohen Wert mit dem zweiten Faktor auf, wird jedoch beim dritten Faktor nicht negativ. Dieses spiegelt die Modellidee wider, dass Personen mit überwiegender FMa eine hohe Hoffnung auf Erfolg haben, aber zudem auch eine hohe Angst vor Misserfolg. Durch diese Faktorenanalyse wird ebenfalls ersichtlich, dass sich die Motivkomponenten bis auf die FM des RLMI-K/J in das quadripolare Modell der Leistungsmotivation einbinden lassen.

Um dieses Ergebnis noch genauer zu untersuchen, um die zweite Hypothese zu prüfen und um zu berechnen, wie sich die einzelnen Leistungsmotive im Raum verteilen bzw. ob diese durch die beiden unabhängigen Dimensionen FM und HE darstellbar sind, wurde eine multidimensionale Skalierung vorgenommen. Zunächst wurden die Werte standardisiert, um vergleichbare Werte zwischen den verschiedenen Fragebögen und deren Konstrukten zu erhalten. Im Anschluss daran wurde eine multidimensionale Skalierung (MDS-Alscal) berechnet. Bei der MDS werden Positionierungsanalysen durchgeführt. Dabei werden ausschließlich globale Ähnlichkeiten, die zwischen den Objekten wahrgenommen werden – hier die einzelnen Leistungsmotive – erfragt (Backhaus et al., 2008). Durch diese berechneten Ähnlichkeiten können dann mithilfe der MDS die Wahrnehmungsdimensionen, hier FM und HE, die zugrunde liegen, abgeleitet werden. Anschließend können die Objekte – die Leistungsmotivkomponenten – mittels dieser Dimensionen im Raum positioniert und grafisch dargestellt werden (Backhaus et al., 2008).

Die multidimensionale Skalierung wurde sowohl mit der FM des RLMI-K/J als auch ohne diese Motivkomponente durchgeführt. Für das Modell mit diesem Motiv wurde ein Stress-Wert (nach Kruskal) von .02 berechnet, was einem ausreichenden Wert entspricht sowie ein akzeptabler Wert von .97 für R-Quadrat (R^2). Das Stress-Maß gibt dabei an, wie die Disparitäten optimal an die Distanzen angepasst sind und R^2 stellt die quadrierte Korrelation der Distanzen mit den Disparitäten dar. Zudem ist es auch ein Maß für die lineare Anpassung dieser Disparitäten an die Varianz bzw. wie viel Varianz der Disparitäten durch die Distanzen erklärt wird. Berechnet man das Modell nun ohne

die FM des RLMI-K/J, ergeben sich weitaus bessere Werte für das Modell. Der Stress-Wert von .00 liegt deutlich unter dem Wert von .10, weshalb die Anpassung als hervorragend bezeichnet werden kann. Aber auch das R^2 von .99 ist ein höchst zufriedenstellender Wert. Das Euklidische Distanzmodell für die Berechnung ohne die Motivkomponente FM des RLMI-K/J zeigt die Abbildung 4.1.

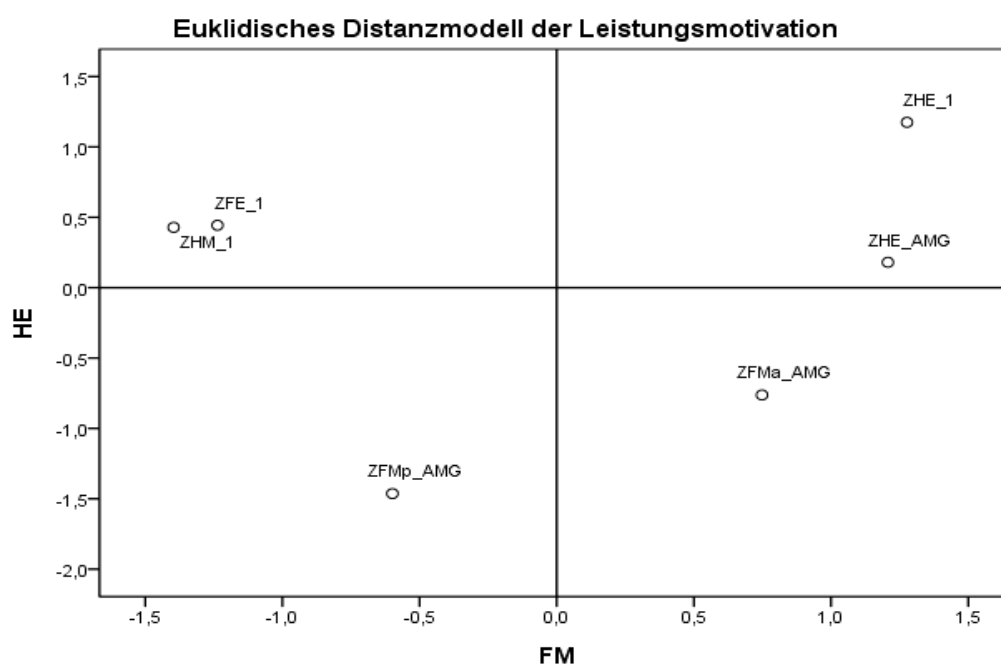


Abbildung 4.1

Euklidisches Distanzmodell für die Motive des AMG-S K-J und des RLMI-K/J ohne FM

Wie aus dieser Abbildung ersichtlich wird, verhalten sich die verbleibenden Motive entsprechend dem quadripolaren Modell der Leistungsmotivation von Covington et al. (Covington & Omelich, 1991; Covington & Roberts, 1994). So verteilen sie sich auf vier Ebenen und lassen sich zudem durch die zwei unabhängigen Dimensionen HE und FM darstellen. Darüber hinaus liefert das Modell mit zwei Dimensionen bessere Werte (siehe oben) als das Modell mit nur einer Dimension (Stress-Wert = .20; R^2 = .88). Der Abbildung ist ebenfalls zu entnehmen, dass sich die Motive FE und HM des RLMI-K/J sehr ähnlich verhalten, da diese nah beieinanderliegen. Konstrukte, die nicht sehr ähnlich zueinander sind, wie beispielsweise HE und FMp, liegen weit voneinander entfernt. Die Distanz zwischen FMa und HE des AMG-S K-J ist relativ gering, da die FMa auch durch eine hohe Hoffnung auf Erfolg gekennzeichnet ist.

Festzuhalten bleibt demnach, dass sich die Leistungsmotivation durch das quadripolare Modell darstellen lässt, also bestehend aus vier unterschiedlichen Komponenten,

mit einer unterschiedlichen Ausprägung von HE und FM. Bei der Einbindung in die vier Komponenten – LL, HH, HL und LH – kann jedoch das Motiv FM des RLMI-K/J nicht berücksichtigt werden, deshalb wird die Hypothese 1 abgelehnt. Da die Einbindung der verbleibenden Motive in das Modell möglich war, welches von zwei unabhängigen Dimensionen HE und FM ausgeht, kann angenommen werden, dass die zweite Hypothese bestätigt wurde. Ob sich die vier Komponenten im Hinblick auf die beiden Dimensionen entsprechend der Abbildung 2.1 (Kapitel 2.1) verhalten, wird im Anschluss an die Clusterverteilung geklärt.

Die weiteren Berechnungen der Leistungsmotive erfolgten mit den vier Konstrukten des quadripolaren Modells, wobei das Motiv FM des RLMI-K/J nicht mehr berücksichtigt wurde. Um die Versuchspersonen den entsprechenden Konstrukten zuordnen zu können, wurden die standardisierten Werte der Motive des RLMI-K/J und des AMG-S K-J einer Clusteranalyse unterzogen und auf vier Cluster verteilt.

Um diese Cluster entsprechend des Euklidischen Distanzmodells bilden zu können, wurden zunächst die standardisierten Werte der Leistungsmotive HE des RLMI-K/J und des AMG-S K-J sowie die Motive FE und HM des RLMI-K/J miteinander addiert und wieder standardisiert. Im Anschluss daran konnten die Versuchspersonen gemäß den berechneten Werten in die verschiedenen Cluster eingeteilt werden. Sie wurden demjenigen Cluster mit der höchsten Ausprägung zugeordnet, wobei jede Versuchsperson auch Werte in den anderen Leistungsmotiven aufweist. Entsprechend dieser Aufteilung konnten 78 Versuchspersonen dem Leistungsmotivtyp LH, 74 Versuchspersonen dem Motiv LL, 91 Versuchspersonen dem Motiv HL und 70 Versuchspersonen dem Leistungsmotiv HH zugeordnet werden.

Der Abbildung 4.2 kann entnommen werden, wie sich die Werte der Leistungsmotive in den verschiedenen Clustern, standardisiert und nicht standardisiert, verteilen. Werden nun die verschiedenen Werte betrachtet, so fällt auf, dass entsprechend der Einteilung das Motiv HH bei Z-FMa den höchsten Wert aufweist. Zudem zeigt sich hier ein positiver Wert beim standardisierten Leistungsmotiv HE des AMG-S K-J, was dem Modell entspricht; denn Personen mit einer hohen FMa haben auch eine hohe Hoffnung auf Erfolg. Das Cluster HL, das aus den beiden Motiven HE des RLMI-K/J und des AMG-S K-J gebildet wird, weist bei diesen die höchsten Werte auf. Auch das Cluster LL zeigt die höchsten Werte bei FE und HM des RLMI-K/J, genau wie das Cluster LH bei FMp des AMG-S K-J.

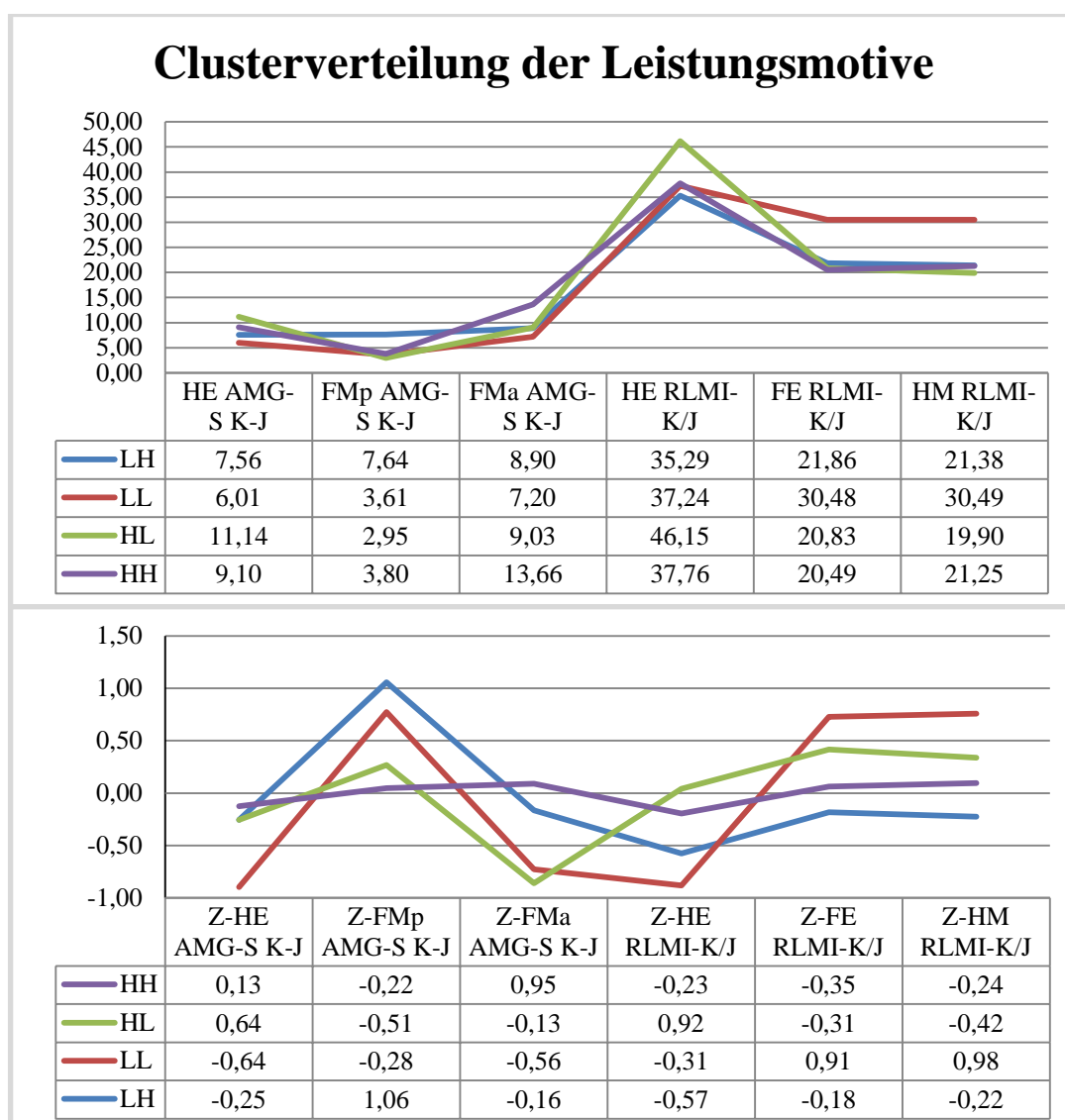


Abbildung 4.2

Clusterverteilung der Leistungsmotive

Die Ergebnisse zeigen, dass eine Einteilung der verschiedenen Leistungsmotive in die vier Cluster – also die Einbindung in das quadripolare Modell – möglich ist. Mithilfe der standardisierten Clustereinteilung wurden die weiteren Berechnungen des Leistungsmotivs vorgenommen.

Doch zunächst wurde noch überprüft, ob sich die vier Cluster hinsichtlich der beiden Dimensionen FM und HE entsprechend der Abbildung 2.1 (siehe Kapitel 2.1) verhalten. Für diese Zwecke wurde mit den vier Motiven eine Faktorenanalyse berechnet. Die vier Komponenten lassen sich durch zwei Faktoren HE und FM mit einer erklärten Gesamtvarianz von 65.35 % darstellen. Die rotierte Komponentenmatrix ist in der Tabelle 4.3 dargestellt.

Tabelle 4.3

Komponentenmatrix der vier Leistungsmotive

	1	2
HL	.82	-.25
HH	.85	.23
LH	.14	.80
LL	-.13	.67

Gemäß Tabelle 4.3 verhalten sich die Motive HL und HH entsprechend der in Abbildung 2.1 erwarteten Zusammenhänge. So stellt Faktor 1 die unabhängige Dimension HE dar und Faktor 2 die Dimension FM. Das Motiv HL lädt auf dem ersten Faktor positiv und auf dem zweiten negativ, das heißt die HE ist stark ausgeprägt und die FM niedrig. Bei HH sind beide Faktoren positiv, da es in beiden Dimensionen Ausprägungen gibt. Das Motiv LH zeigt zwar eine hohe positive Ladung auf den Faktor 2, sollte aber auf dem ersten Faktor negativ laden, da HE niedrig ausgeprägt sein sollte. Auch das Motiv LL stimmt zumindest in einer Dimension mit den erwarteten Zusammenhängen überein, da es im ersten Faktor negativ ist und dies auf eine geringe HE hinweist. Hier sollte jedoch auch die Ladung des zweiten Faktors negativ sein. So müssen schließlich die erwarteten Zusammenhänge zwischen den unabhängigen Dimensionen und den vier Motiven teilweise zurückgewiesen werden. Aber auch die Faktorenanalyse bestätigt die zwei unabhängigen Dimensionen HE und FM, da hier sowohl die vier Motive unterschiedlich stark laden als auch durch diese erklärt werden können.

4.2 Testmotivation

Zur Überprüfung der Hypothese 3 wurden die verschiedenen Skalen der Testmotivation, die vor und nach dem Test erhoben wurden, analysiert.

H3: Die Testmotivation kann mit den verschiedenen Subskalen der Theorie von Eccles et al. (1983) erfasst werden.

In Kapitel 3.4.5.2 wurden die Skalen der Testmotivation bereits hinsichtlich ihrer Korrelation untersucht, wobei alle Skalen auf dem .01 Niveau signifikant mit der gesamten Testmotivation korrelieren.

Das Modell, auf welchem die Testmotivation basiert, wurde ebenfalls in Kapitel 3.4.5.3 einer konfirmatorischen Faktorenanalyse unterzogen. Dabei ergibt sich durch die Berechnung mithilfe von AMOS für das Modell der Testmotivation mit den Skalen – AWT1, AWT2, EWT1, EWT2, ALV, Einst., LV, wobei die Oberskalen AW, EW und EL gebildet werden – ein χ^2 von 102.62 mit $df=11$ und $p = .00$. Auch die beiden weiteren Indizes der CFI und der RMSEA liegen mit .873 und .159 unter- bzw. oberhalb der kritischen Grenzen, weshalb insgesamt nicht von einem adäquaten Modell-Fit ausgegangen werden kann und das Modell bzw. die Hypothese 3 abgelehnt werden muss.

Da in dem eben genannten Modell die Skalen als beobachtbare Variablen dienen, ist es interessant zu überprüfen, ob sich die Modellgültigkeit verbessert, wenn anstelle der Skalen die einzelnen Items aus dem Fragebogen als beobachtbare Variablen verwendet werden. In diesem Fall stellen bereits die Skalen AWT1, AWT2, EWT1, EWT2, ALV, Einst. und LV latente Variablen dar, die durch die Items beeinflusst werden und ihrerseits die Oberskalen AW, EW und EL beeinflussen, aus welchen die Testmotivation schließlich gebildet wird. Zu diesem Zweck wurden alle Items in AMOS eingegeben und die Verbindung zu den latenten Variablen hergestellt. Das dabei resultierende Modell kann der Abbildung 4.3 entnommen werden. Aus Gründen der Übersichtlichkeit wurde jedoch auf die Darstellung der einzelnen Items verzichtet.

Für das in der Abbildung 4.3 dargestellte Modell ergibt sich ein χ^2 von 3234.6 mit $df = 1.315$ ohne Angabe von p , da die Lösung, aufgrund von negativen Messfehlern bei AW, EWT2 und Einst., nicht akzeptabel ist. Dabei können die negativen Messfehler durch die überhöhten Korrelationen von Testmotivation zu AW (1.04), EL zu Einst. (1.25) und EW zu EWT2 (1.16) erklärt werden. Für den RMSEA liegt der berechnete Wert von .073 mit 90 % Wahrscheinlichkeit im Intervall [.070; .076], womit alle drei Werte unter der kritischen Grenze von .08 bleiben. Der CFI ist mit .647 jedoch nicht über der Grenze von .90.

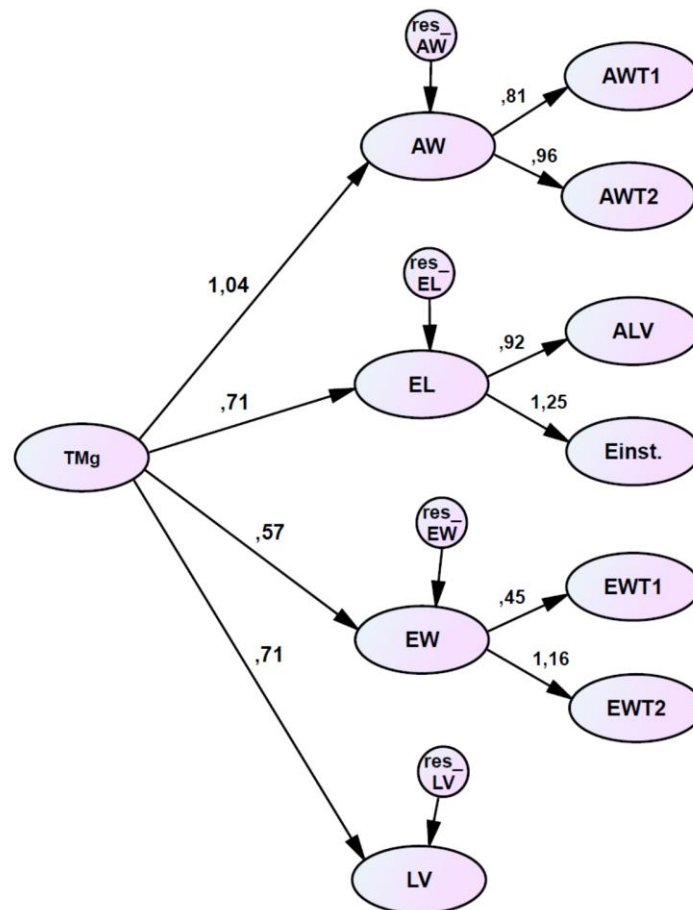


Abbildung 4.3

Modell der Testmotivation mit Items als beobachtbare Variablen

Insgesamt kann allerdings auch hier nicht von einem adäquaten Modell-Fit ausgegangen werden. Weshalb auch für diese Modellrechnung die Hypothese 3 abgelehnt werden muss.

Da die Testmotivation sowohl vor (TMv) als auch nach dem Test (TMn) erhoben worden ist, wurde anschließend überprüft, ob der nicht adäquate Modell-Fit auf die Unterschiede zwischen der Testmotivation vorher bzw. nachher zurückzuführen ist. Aus diesem Grund wurden diese mithilfe der Modellrechnung jeweils mit Skalen bzw. Items als beobachtbare Variablen berechnet. Die Ergebnisse können der Tabelle 4.4 entnommen werden.

Tabelle 4.4

Modellrechnung jeweils mit den Skalen bzw. mit Items der Testmotivation vor bzw. nach dem Test

	TMv Skalen	TMn Skalen	TMv Items	TMn Items
χ^2 ; (df); p	7.71; (2); .02	0.00; (0);	1507.30; (430); .00	763.71; (206); .00
RMSEA	.093; [.031; .166]	-	.087; [.082; .092]	.090; [.084; .097]
CFI	.977	-	.670	.775

Anmerkungen. χ^2 = Testwert; df = Freiheitsgrade; p = Signifikanzwert

In der Tabelle 4.4 wird ersichtlich, dass sich die Modellrechnungen für die Testmotivation getrennt – nach vor und nach dem Test – etwas verbessern, diese jedoch ebenfalls nicht für einen adäquaten Modell-Fit sprechen. Die Modellrechnung für die Testmotivation nach dem Test mit Skalen als beobachtbare Variablen konnte nicht berechnet werden, da sich für diese ein Freiheitsgrad von Null ergab und somit das Modell komplett durch die beobachtbaren Variablen erklärt wird.

Zum Schluss wurde noch der Zusammenhang zwischen der Testmotivation vor und nach dem Test berechnet. Das dabei resultierende Modell ist in der Abbildung 4.4 dargestellt.

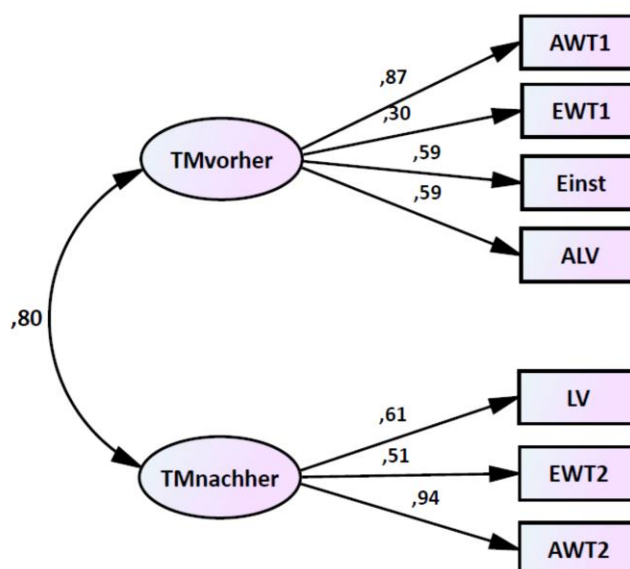


Abbildung 4.4

Modell über den Zusammenhang zwischen der Testmotivation vor und nach dem Test

Entsprechend der Abbildung 4.4 korrelieren die beiden Testmotivationen mit .80. Für dieses Modell ergibt sich ein χ^2 von 94.93 mit $df = 13$ und p von .00. Für den RMSEA liegt der berechnete Wert von .138 mit 90 % Wahrscheinlichkeit im Intervall [.113; .165], womit alle drei Werte über die kritische Grenze von .08 steigen; auch der CFI liegt mit einem Wert von .887 unterhalb der Grenze von .90. Aus diesen Gründen muss auch dieses Modell als inadäquat zurückgewiesen werden.

Zusammenfassend bleibt festzuhalten, dass die Testmotivation mit den verschiedenen Subskalen der Theorie von Eccles et al. (1983) mithilfe der Modellrechnung von AMOS nicht adäquat erfasst werden kann. Werden hingegen ausschließlich die einzelnen Skalen betrachtet (siehe Kapitel 3.4.5.2 und 3.4.5.3), so liefern diese akzeptable bis gute Werte für die interne Konsistenz nach Cronbach's α . Aus diesem Grund wird davon ausgegangen, dass jede Skala für sich genommen einen Teil der Testmotivation erfasst, nur die Zusammenfassung zu einem einheitlichen Modell ist nicht möglich.

4.3 Intelligenz

Bereits in Kapitel 3.4.6.2 wurden die einzelnen Subtests des LPS-neu hinsichtlich verschiedener Parameter sowohl für Version A als auch für Version B untersucht. Entsprechend der dort vorgenommenen Analysen und gewonnenen Ergebnisse wurden die beiden Versionen nachfolgend getrennt voneinander untersucht.

H4: Die formulierte Struktur des LPS-neu von Kreuzpointner (2010) lässt sich replizieren.

Um diese Hypothese zu überprüfen, wurden zunächst alle Subtests des LPS-neu der Versionen A und B mit den Stratum II Komponenten der Intelligenz – kristalline Intelligenz (LPS_Gc), fluide Intelligenz (LPS_Gf), visuelle Wahrnehmung (LPS_GV), kognitive Schnelligkeit (LPS_GS) – des von Kreuzpointner (2010) vorgeschlagenen Modells korreliert (siehe Tabelle 4.5). Entsprechend der formulierten Struktur von Kreuzpointner (2010) sollten die Subtests 1 und 2 mit der kristallinen Intelligenz, die Subtests 3 bis 5 mit der fluiden Intelligenz, die Subtests 6, 7 und 8 mit der visuellen Wahrnehmung sowie die Subtests 9 bis 11 mit der kognitiven Schnelligkeit am höchsten korrelieren.

Tabelle 4.5

Korrelationen der Subtests mit den einzelnen Dimensionen des LPS-neu, getrennt für Version A und B

A	LPS_Gc	LPS_Gf	LPS_Gv	LPS_Gs	B	LPS_Gc	LPS_Gf	LPS_Gv	LPS_Gs
1	.86**	.42**	.33**	.31**	1	.83**	.26**	.35**	.19*
2	.87**	.42**	.26**	.36**	2	.86**	.36**	.33**	.13
3	.38**	.74**	.39**	.30**	3	.23**	.75**	.44**	.28**
4	.30**	.78**	.23**	.23**	4	.27**	.74**	.33**	.27**
5	.47**	.80**	.42**	.31**	5	.35**	.77**	.35**	.29**
6	.27**	.23**	.77**	.11	6	.31**	.34**	.74**	.34**
7	.20**	.39**	.66**	.29**	7	.21**	.36**	.66**	.26**
8	.25**	.36**	.70**	.25**	8	.35**	.41**	.81**	.18*
9	.23**	.25**	.19*	.60**	9	.11	.22**	.22**	.57**
10	.26**	.20**	.21**	.82**	10	.14	.30**	.30**	.90**
11	.32**	.34**	.19*	.60**	11	.14	.29**	.20**	.45**

Anmerkungen. **) Die Korrelation ist auf dem .01 Niveau signifikant; *) Die Korrelation ist auf dem .05 Niveau signifikant

Die höchsten Zusammenhänge der Subtests mit den Stratum II Komponenten der Intelligenz sind in der Tabelle 4.5 grau hinterlegt.

Um die Struktur genauer untersuchen zu können, wurde sie mithilfe von AMOS einer Strukturgleichungsmodellierung unterzogen. Für diese Zwecke wurden zunächst die Mahalanobis Distanzen für die beiden Versionen getrennt berechnet und nach Versuchspersonen gesucht, deren Leistungen irregulär von der durchschnittlichen Leistung im LPS-neu abweichen. Hierbei fielen vor allem zwei Versuchspersonen mit einem Abstand von 45.18 (Version A) und 34.36 (Version B) auf, bei einem durchschnittlichen Abstand von 10.94 ($SD = 5.50$ der Version A; $SD = 6.42$ der Version B). Beide Versuchspersonen weisen in zwei der elf Subtests besonders niedrige Werte auf, jedoch in den verbleibenden Subtests sehr hohe Werte. Es kann also nicht darauf geschlossen werden, dass sie für die Bearbeitung nicht ausreichend motiviert waren, sondern viel eher, dass ihnen die neun Subtests mehr entsprachen als die zwei Subtests, in denen sie schlechter abschnitten. Aus diesem Grund wurde keine Versuchsperson aus den weiteren Berechnungen ausgeschlossen. So konnten für die Modellrechnung der Version A des LPS-neu 178 Versuchspersonen und für die Version B 183 Versuchspersonen berücksichtigt werden.

Der folgenden Abbildung 4.5 können die einzelnen Zusammenhänge zwischen der allgemeinen Intelligenz und Stratum II sowie die Zusammenhänge zwischen Stratum II mit den einzelnen Subtests sowohl für die Version A (erstes Modell) als auch für die Version B (zweites Modell) entnommen werden. Aus den dargestellten Modellen wird ersichtlich, dass die fluide Intelligenz bei beiden Versionen (A: .95; B: .92) am höchsten mit dem Gesamtwert des LPS-neu korreliert. Der geringste Zusammenhang mit dem LPS-neu Gesamtwert zeigt sich bei der Version A in der Dimension visuelle Wahrnehmung (.77) und bei Version B in der kristallinen Intelligenz (.67).

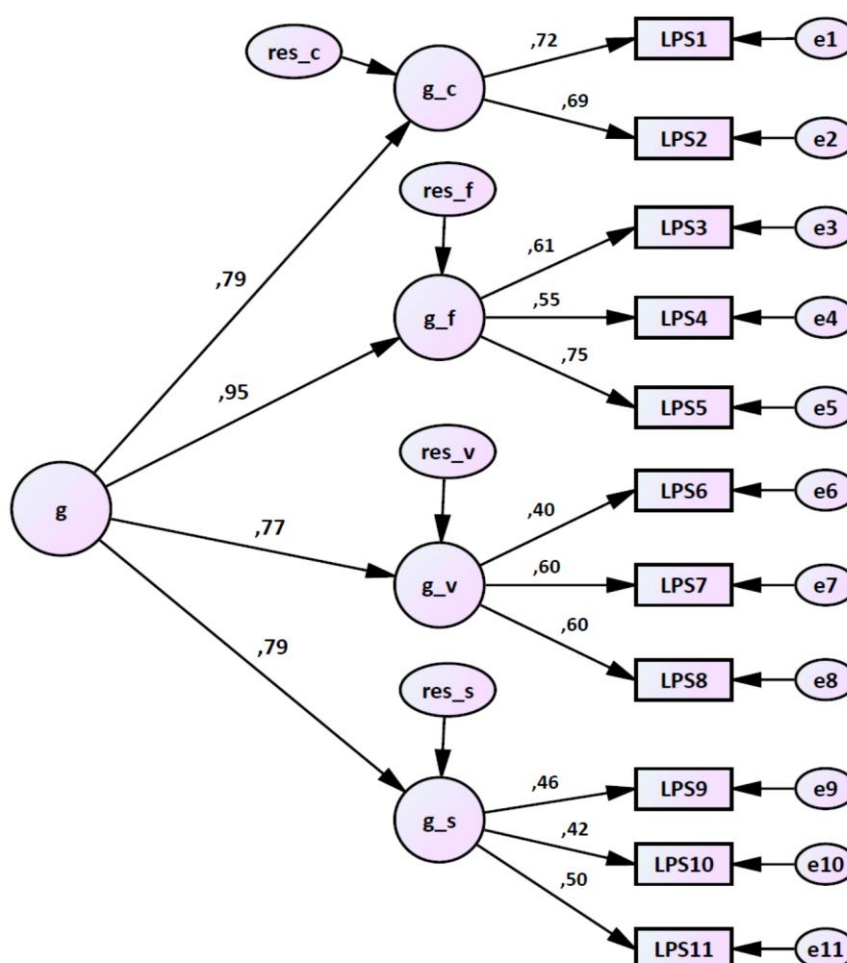
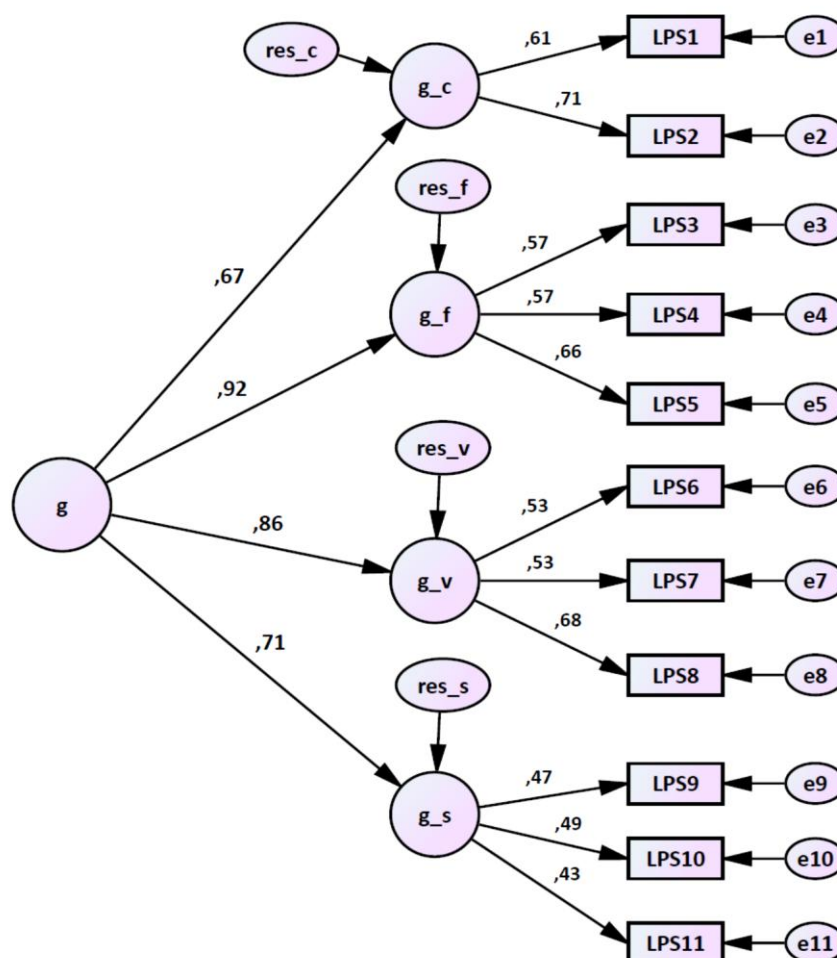


Abbildung 4.5

Strukturmodell des LPS-neu getrennt für Version A und B



Fortsetzung: Abbildung 4.5

Strukturmodell des LPS-neu getrennt für Version A und B

Werden diese Modelle nun berechnet und nur die Version A in den Modellrechnungen des LPS-neu berücksichtigt, ergibt sich ein χ^2 von 46.79 mit $df = 40$ und $p = .22$. Für den RMSEA liegt der berechnete Wert von .031 mit 90 % Wahrscheinlichkeit im Intervall [.000; .063]. Somit bleibt dieser Wert unter der kritischen Grenze von .08. Auch der CFI ist mit .979 über dem kritischen Wert von .90. Alle drei Koeffizienten sprechen für die Modellgültigkeit des LPS-neu der Version A. Die Ergebnisse für die Version B des LPS-neu fallen hingegen etwas schlechter aus. So ergibt sich hier ein χ^2 von 60.82 mit $df = 40$ und $p = .02$. Für den RMSEA liegt der berechnete Wert von .053 mit 90 % Wahrscheinlichkeit im Intervall [.022; .079]. Zusammen mit einem CFI von .933 weisen trotzdem beide Indizes auf die Modellgültigkeit der Version B des LPS-neu hin.

Zusammenfassend kann festgehalten werden, dass sich die Struktur des LPS-neu, für beide Versionen getrennt, replizieren lässt. Dementsprechend kann davon ausgegangen werden, dass die Struktur des LPS-neu gültig ist und die Hypothese 4 als bestätigt gilt.

4.4 Zusammenhänge

In den nachfolgenden Abschnitten sollten nun die Hypothesen und Fragestellungen geprüft werden, die sich mit den Zusammenhängen zwischen den einzelnen Messinstrumenten beschäftigen. Dabei wurde zunächst der Zusammenhang zwischen der Testmotivation und den vier Leistungsmotiven überprüft. Im Anschluss daran wurde der Zusammenhang zwischen Testmotivation und Intelligenz untersucht und zum Abschluss der Zusammenhang zwischen Schulleistung, besuchter Schulart bzw. Jahrgangsstufe und Intelligenz.

4.4.1 Leistungsmotiv und Testmotivation

In diesem Kapitel sollte geklärt werden, welche Leistungsmotive, entsprechend der Hypothese 5, eine ausreichende Testmotivation für die Bearbeitung des Leistungstests aufweisen. Zudem wurden auch die beiden unten stehenden Fragestellungen F1 und F2 beantwortet.

- H5: Versuchspersonen, bei denen das Leistungsmotiv HH, HL bzw. LH hoch ausgeprägt ist, weisen eine ausreichende Testmotivation für die Bearbeitung eines Leistungstests auf.*
- F1: Gibt es Unterschiede, abhängig vom Leistungsmotivtyp, in der Höhe der vorliegenden Testmotivation (HH bzw. HL höher als LH bzw. LL)?*
- F2: Kann bei Kenntnis des Leistungsmotivs auf die Testmotivation geschlossen werden?*

Bevor mit den Berechnungen begonnen werden konnte, mussten die Datensätze zusammengeführt und nach fehlenden Werten analysiert werden. Entsprechend konnten nur Versuchspersonen mit vollständigen Datensätzen in allen drei Fragebögen berücksichtigt werden. Wie bereits in den Kapiteln 3.4.3 und 3.4.5 beschrieben, wurden fehlende Items, sofern nur eines pro Skala fehlte, durch Mittelwerte ersetzt. Aus diesem Grund konnten für die Berechnungen über die Zusammenhänge zwischen den Leistungsmotivtypen und der Testmotivation insgesamt 293 Versuchspersonen berücksichtigt werden. Für die in Kapitel 4.1 berechneten Cluster der Leistungsmotivtypen verblieben somit 68 Versuchspersonen für das Leistungsmotiv LH, 73 für LL, 83 für HL und 69 für das Leistungsmotiv HH.

Zunächst wurde berechnet, ob sich die vier Leistungsmotivcluster augenscheinlich in der Höhe ihrer gesamten Testmotivation bzw. in der Testmotivation vor und auch nach dem Test unterscheiden. Vergleicht man nun die Mittelwerte der einzelnen Motive, so fällt auf, dass sie sich teilweise deutlich voneinander unterscheiden (siehe Tabelle 4.6).

Tabelle 4.6

Testmotivationsmittelwerte und Standardabweichungen der Leistungsmotivcluster

	TMg <i>M (SD)</i>	TMg <i>Min</i>	TMg <i>Max</i>	TMv <i>M (SD)</i>	TMv <i>Min</i>	TMv <i>Max</i>	TMn <i>M (SD)</i>	TMn <i>Min</i>	TMn <i>Max</i>
LH	157.90 (23.55)	81.00	211.00	92.49 (14.96)	47.00	135.00	65.42 (11.49)	34.00	92.00
LL	162.56 (25.71)	104.96	209.00	95.93 (15.66)	60.00	122.00	67.03 (12.99)	41.00	93.00
HL	171.31 (25.05)	114.00	221.00	100.77 (15.22)	67.00	129.00	70.54 (13.52)	32.00	95.00
HH	173.42 (24.26)	116.00	232.00	99.99 (13.58)	67.00	135.00	73.43 (12.56)	37.00	103.00

Anmerkungen. *M* = Mittelwert; *SD* = Standardabweichung

Die Ergebnisse zeigen, dass die Leistungsmotive, die von einer hohen Hoffnung auf Erfolg geprägt sind, auch eine höhere Testmotivation aufweisen, sowohl insgesamt als auch vor und nach dem Test. Die höchste Testmotivation zeigt vor allem die Gruppe, die zusätzlich noch eine hohe Furcht vor Misserfolg besitzt (HH), also in beiden Dimensionen hohe Werte aufweist. Die geringste Testmotivation weist entgegen der Hypothese 5 die Gruppe LH auf, bei der die Furcht vor Misserfolg überwiegt und nicht, wie vermutet, das Cluster LL, wo sowohl die Furcht vor Misserfolg als auch die Hoffnung auf Erfolg gering ausgeprägt sein sollte. Zur Veranschaulichung sind die Mittelwerte der gesamten Testmotivation nochmals als Grafik in der Abbildung 4.6 dargestellt.

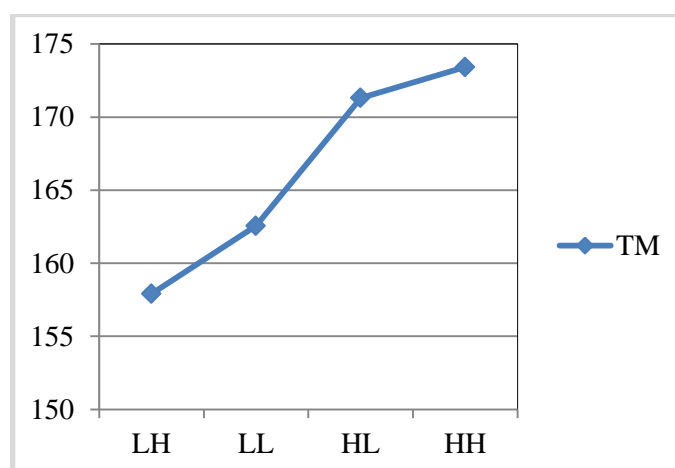


Abbildung 4.6

Testmotivationsmittelwerte der Leistungsmotivtypen

Doch unterscheiden sich die vier Cluster in ihrer Testmotivation statistisch signifikant voneinander?

Zur Beantwortung dieser Frage wurden die vier Cluster einer univariaten Varianzanalyse unterzogen. Diese ergab für die gesamte Testmotivation ein signifikantes Ergebnis ($F(3) = 6.23$; $p = .00$; $\eta^2 = .07$); auch für die Testmotivation vorher ($F(3) = 4.96$; $p = .00$; $\eta^2 = .05$) und nachher ($F(3) = 5.57$; $p = .00$; $\eta^2 = .06$) werden die Ergebnisse signifikant. Es wird deutlich, dass sich die Gruppen signifikant voneinander unterscheiden. Bei Betrachtung des Scheffé Post-Hoc-Tests wird klar, welche Leistungsmotivcluster sich in der Ausprägung ihrer Testmotivation signifikant unterscheiden (siehe Tabelle 4.7).

Die Gruppe des Leistungsmotivtyps LH unterscheidet sich in ihrer Testmotivation sowohl insgesamt als auch in der Testmotivation nach dem Test signifikant in ihrer Höhe von HH und HL, jedoch nicht von LL (siehe Tabelle 4.7). Für die Testmotivation, die nach dem Test erhoben wurde, unterscheidet sich das Cluster LH hingegen nur von HH signifikant. Bei Betrachtung der gesamten Testmotivation fällt auf, dass der Leistungsmotivtyp LL zu einem in die Gruppe von HL und HH und zum anderen auch in die Gruppe von LH passt.

Tabelle 4.7

Mittelwertsdifferenzen und Signifikanzen zwischen den Leistungsmotivtypen in Abhängigkeit von der Testmotivation

TMg		Mitt. Diff.	Stand. fehler	p	TMv		Mitt. Diff.	Stand. fehler	p	TMn		Mitt. Diff.	Stand. fehler	p
LH	LL	-4.65	4.16	.74	LH	LL	-3.04	2.51	.69	LH	LL	-1.61	2.14	.90
	HL	-13.41	4.04	.01		HL	-8.29	2.44	.01		HL	-5.12	2.08	.11
	HH	-15.52	4.22	.00		HH	-7.50	2.55	.04		HH	-8.02	2.17	.00
LL	LH	4.65	4.16	.74	LL	LH	3.04	2.51	.69	LL	LH	1.61	2.14	.90
	HL	-8.75	3.96	.18		HL	-5.24	2.39	.19		HL	-3.51	2.04	.40
	HH	-10.87	4.15	.08		HH	-4.46	2.50	.37		HH	-6.40	2.13	.03
HL	LH	13.41	4.04	.01	HL	LH	8.29	2.44	.01	HL	LH	5.12	2.08	.11
	LL	8.75	3.96	.18		LL	5.24	2.39	.19		LL	3.51	2.04	.40
	HH	-2.11	4.02	.96		HH	.78	2.42	.99		HH	-2.89	2.07	.58
HH	LH	15.52	4.22	.00	HH	LH	7.50	2.55	.04	HH	LH	8.02	2.17	.00
	LL	10.87	4.15	.08		LL	4.46	2.50	.37		LL	6.40	2.13	.03
	HL	2.11	4.02	.96		HL	-7.78	2.42	.99		HL	2.89	2.07	.58

Anmerkung. p = Signifikanzwert

Somit können die beiden oben genannten Fragestellungen beantwortet werden: Die vier Leistungsmotivtypen unterscheiden sich signifikant in ihrer Höhe der Testmotivation voneinander, wobei die Versuchspersonen, bei denen das Motiv Hoffnung auf Erfolg hoch ausgeprägt ist, auch eine höhere Testmotivation aufweisen. Die höchste Testmotivation weist das Motiv HH auf, gefolgt von HL und LL, und die geringste Testmotivation weist das Motiv LH auf. Zwar unterscheiden sich die vier Leistungsmotivtypen im Mittel signifikant in den Werten ihrer Testmotivation, betrachtet man aber die Minimum- und Maximumwerte der einzelnen Gruppen (siehe Tabelle 4.6), so fallen die Maximumwerte zwar bei den Leistungsmotivtypen LL und LH etwas niedriger aus, aber auch die Minimumwerte der beiden anderen Motive weisen teilweise geringe Werte auf. Aus diesem Grund lässt sich bei Kenntnis des Leistungsmotivtyps nicht eindeutig auf die Testmotivation schließen.

Um nun die Hypothese 5 beantworten zu können, musste zunächst geklärt werden, wann eine ausreichende Testmotivation vorliegt und ob sich die jeweiligen Leistungsmotivtypen von diesem Trennwert signifikant unterscheiden. Für die Einteilung wird die Testmotivation in drei Gruppen aufgeteilt, wobei die unteren 25 % nicht ausreichend testmotiviert, die mittleren 50 % ausreichend testmotiviert und die oberen 25 % sehr testmotiviert sind. Gemäß dieser Einteilung besteht die Gruppe der ausreichend Testmo-

tivierten aus 149 Versuchspersonen, die einen Mittelwert von 167.02 ($SD = 9.53$) in der gesamten Testmotivation aufweisen (TMv: $M = 97.00$ ($SD = 8.07$); TMn: $M = 70.01$ ($SD = 7.97$). Dieser Wert wurde nun als Trennwert für eine ausreichende Testmotivation verwendet und mithilfe eines t -Testes mit den Mittelwerten der vier Cluster auf Signifikanz untersucht. Die Ergebnisse können der Tabelle 4.8 entnommen werden.

Tabelle 4.8

Mittelwertsvergleich der verschiedenen Leistungsmotivcluster mit ausreichender mittlerer Testmotivation

TMg	t (df)	p	TMv	t (df)	p	TMn	t (df)	p
LH	-3.19 (67)	.00	LH	-2.49 (67)	.02	LH	-3.30 (67)	.00
LL	-1.48 (72)	.14	LL	-.80 (72)	.43	LL	-1.96 (72)	.05
HL	1.56 (.82)	.12	HL	2.26 (82)	.03	HL	.36 (82)	.72
HH	2.19 (68)	.03	HH	1.83 (68)	.07	HH	2.26 (68)	.03

Anmerkungen. t = Testwert; df = Freiheitsgrade; p = Signifikanzwert

Wie Tabelle 4.8 zeigt, liegt nur bei dem Leistungsmotivtyp LH eine signifikant nicht ausreichende Testmotivation – insgesamt, vor und nach dem Test – vor. Bei allen anderen Leistungsmotiven liegt entweder eine ausreichende oder sogar eine über dem Mittelwert der ausreichenden Testmotivation liegende Motivation für die Bearbeitung des Tests vor.

Es bleibt demnach festzuhalten, dass entgegen der Hypothese 5 nicht das Leistungsmotiv LL, sondern das Motiv LH keine ausreichende Testmotivation besitzt. HH und HL verhalten sich hingegen entsprechend der Hypothese und besitzen eine ausreichende Motivation für die Testung. Somit muss die Hypothese zum jetzigen Zeitpunkt abgelehnt werden.

4.4.2 Testmotivation und Leistungen in einem Leistungstest

In diesem Abschnitt sollte der Zusammenhang zwischen der Testmotivation und dem Leistungsprüfungs-system-neu überprüft werden. Dabei wurde zunächst in Hypothese 6 untersucht, ob die Testmotivation einen Einfluss auf die Leistungen in diesem Leistungstest ausübt. In den weiteren Hypothesen wurde dann dieser Zusammenhang entsprechend der Hypothesen 7 und 8 näher untersucht. Doch zunächst folgten die Berechnungen zur anschließenden Hypothese.

H6: Die Testmotivation hat einen positiven Einfluss auf die Leistungen des LPS-neu.

Wie bereits im Kapitel 3.4.5.3 dargestellt, konnten für die Berechnungen der Testmotivation lediglich 332 Versuchspersonen berücksichtigt werden. Um zu überprüfen, ob und gegebenenfalls wie stark die Testmotivation die Leistungen in der revidierten Version des Leistungsprüfsystems beeinflusst, wurden zunächst bivariate Korrelationen nach Bravais-Pearson zwischen diesen beiden Messinstrumenten berechnet (siehe Tabelle 4.9). Dabei wurden die Versionen A und B für die Berechnungen der Stratum II Komponenten des LPS-neu getrennt und entsprechend der in Kapitel 3.4.6.2 gefundenen Äquivalenz bei Betrachtung des LPS-neu Gesamtwertes zusammen berechnet. Die Ergebnisse dieser Berechnungen können der Tabelle 4.9 entnommen werden.

Tabelle 4.9

Korrelationen zwischen den Skalen der Testmotivation und den Stratum II Komponenten der Intelligenz sowie dem Gesamtwert des LPS-neu

A	Gc	Gf	Gv	Gs	B	Gc	Gf	Gv	Gs	A und B	LPS_G
TMg	.03	.19*	.13	.15	TMg	.10	.17*	.15	.05	TMg	.16**
TMv	-.02	.10	.08	.09	TMv	.10	.07	.09	-.06	TMv	.07
TMn	.09	.26**	.17*	.19*	TMn	.08	.24**	.18*	.16*	TMn	.23**
AWT1	.02	.11	.13	.07	AWT1	.08	.11	.10	-.02	AWT1	.10
AWT2	.04	.15	.13	.11	AWT2	.03	.17*	.16*	.10	AWT2	.15**
AWG	.03	.14	.14	.10	AWG	.06	.15*	.14	.04	AWG	.13*
ALV	-.02	.06	.02	.05	ALV	.08	.00	.02	-.05	ALV	.02
Einst.	-.01	.09	.06	.15	Einst.	-.04	-.06	.00	-.11	Einst.	.01
EL	-.02	.08	.04	.10	EL	.05	-.03	.01	-.08	EL	.02
EWT1	-.10	-.01	-.08	.01	EWT1	.16*	.07	.13	-.07	EWT1	.04
EWT2	.03	.23**	.14	.11	EWT2	.19*	.21**	.26**	.15	EWT2	.23**
EWG	-.05	.12	.02	.06	EWG	.22**	.17*	.24**	.05	EWG	.16**
LV	.14	.28**	.15*	.25**	LV	.03	.22**	.09	.17*	LV	.02

Anmerkungen. **) Die Korrelation ist auf dem .01 Niveau signifikant; *) Die Korrelation ist auf dem .05 Niveau signifikant

Die gesamte Testmotivation weist einen signifikanten positiven Zusammenhang ($r = .16$, $p = .00$) mit dem Gesamtwert des LPS-neu auf (siehe Tabelle 4.9). Werden nur die Stratum II Komponenten im Zusammenhang mit der gesamten Testmotivation betrachtet, so zeigt sich die höchste positive Korrelation sowohl bei Version A als auch bei Version B mit der fluiden Intelligenz (A = .19; B = .17). Die Testmotivation, die nach

dem Test erhoben wurde, zeigt, im Gegensatz zur TMv, über fast alle Stratum II Komponenten hinweg und auch beim Gesamtwert des LPS-neu höhere und meist signifikante Zusammenhänge. So reicht die Spannweite der TMn von $r = .08$ für kristalline Intelligenz (bei Version B) bis $r = .26$ für fluide Intelligenz (bei Version A). Bei Betrachtung der einzelnen Skalen der Testmotivation mit dem LPS_G weist die Skala Einstellungen den niedrigsten Zusammenhang ($r = .01$; $p = .84$) und die Erfolgserwartung nach dem Test ($r = .23$; $p = .00$) den höchsten Zusammenhang auf. Ferner lässt sich erkennen, dass die Skalen Einstellungen, allgemeines Leistungsverhalten und Leistungsverhalten sehr geringe oder gar negative Werte mit dem LPS-neu Gesamtwert aufweisen. Zudem zeigen die beiden Skalen Einstellungen und allgemeines Leistungsverhalten für beide Versionen des LPS-neu ebenfalls geringe bis negative Werte mit den Stratum II Komponenten des Leistungsprüfsystems. Insgesamt kann jedoch davon ausgegangen werden, dass die Testmotivation einen geringen aber signifikanten Einfluss auf die Leistungen des Leistungsprüfsystems-neu ausübt. Dieser Zusammenhang wurde nun anhand des in Kapitel 3.4.1 formulierten Strukturgleichungsmodells mithilfe der Modellrechnung von AMOS überprüft. Aufgrund der in Kapitel 3.4.4 vorgenommenen Analysen und der dort errechneten Unterschiede zwischen den Versionen A und B des LPS-neu wurden diese auch in den Berechnungen mit der Testmotivation getrennt voneinander überprüft.

Es ergibt sich dabei für die Berechnung des Modells der Testmotivation mit den Skalen – AWT1, AWT2, EWT1, EWT2, ALV, Einst., LV, wobei die Oberskalen AW, EW und EL gebildet werden – und dem Leistungsprüfsystem-neu der Version A ein χ^2 von 185.79 mit $df = 127$ und $p = .00$. Das entsprechende Modell zeigt die Abbildung 4.7. Gemäß diesen Berechnungen übt die Testmotivation mit .25 einen Einfluss auf die Leistungen des Leistungsprüfsystems aus. Zudem liegt der für den RMSEA berechnete Wert von .053 mit 90 % Wahrscheinlichkeit im Intervall [.035; .068]; damit bleiben alle drei Werte unter der kritischen Grenze von .08. Auch der CFI ist mit .915 über dem kritischen Wert von .90. Beide Koeffizienten sprechen für die Modellgültigkeit des LPS-neu der Version A in Verbindung mit der Testmotivation.

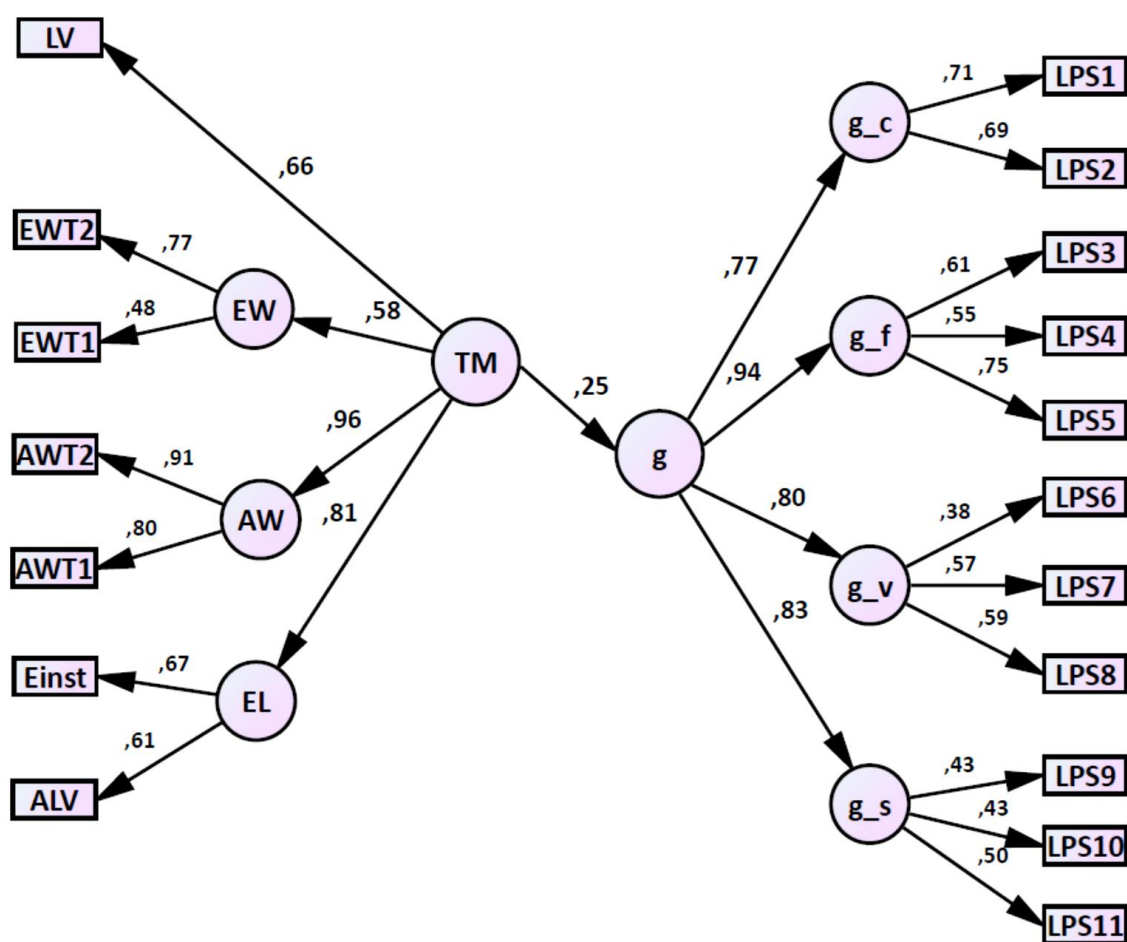


Abbildung 4.7

Modell über den Zusammenhang zwischen der Testmotivation mit Skalen als beobachtbare Variablen und dem LPS-neu (Version A)

Die Ergebnisse für die Version B des LPS-neu (siehe Anhang D-1) fallen hingegen auch hier wieder etwas schlechter aus, wobei die Testmotivation in diesen Berechnungen eine Korrelation von .29 mit dem Gesamtwert des LPS-neu aufweist. So ergibt sich hier ein χ^2 von 218.07 mit $df = 127$ und $p = .02$. Für den RMSEA liegt der berechnete Wert von .066 mit 90 % Wahrscheinlichkeit im Intervall [.051; .081] und der CFI liegt bei .861, weshalb dieses Modell als inadäquat zurückgewiesen werden müsste.

Wie in Kapitel 4.4.1 sollte nun ebenfalls überprüft werden, ob sich die Modellrechnungen zwischen Testmotivation und LPS-neu verbessern, wenn die Testmotivation getrennt untersucht wird. Die Ergebnisse können in der Tabelle 4.10 nachgelesen werden.

Tabelle 4.10

Modellrechnung der Testmotivation vor bzw. nach dem Test mit dem LPS-neu

	TMv Version A	TMv Version B	TMn Version A	TMn Version B
$\chi^2; (df); p$	103.95; (85); .08	115.21; (85); .02	91.57; (72); .06	91.26; (72); .06
RMSEA	.037 [.000;.059]	.047 [.021;.067]	.040 [.000; .063]	.041 [.000;.064]
CFI	.955	.931	.955	.954
r	.14	.08	.37	.35

Anmerkungen. χ^2 = Testwert; df = Freiheitsgrade; p = Signifikanzwert

Tabelle 4.10 zeigt, dass sich die Modellrechnung für die Testmotivation, getrennt nach vor und nach dem Test, deutlich verbessert und die Modelle alle für einen adäquaten Modell-Fit sprechen. Zudem wird ebenfalls ersichtlich, wie bereits der Tabelle 4.9 zu entnehmen ist, dass die Testmotivation nach dem Test höhere Korrelationen mit dem LPS-neu Gesamtwert aufweist als die vor dem Test. Entsprechend der Ergebnisse für die Modellrechnungen lässt sich festhalten, dass der Einfluss der Testmotivation auf die Leistungen im LPS-neu besser erklärt werden kann, wenn die Testmotivation getrennt und ohne Oberskalen betrachtet wird. Aus diesem Grund wurde nun das Modell für die Testmotivation und deren Einfluss auf den Gesamtwert des LPS-neu überprüft, wobei die Testmotivation getrennt nach vor und nach dem Test im Modell integriert wurde.

Das dabei errechnete Modell für die Version A des LPS-neu ist in Abbildung 4.8 dargestellt. Entsprechend dieser Berechnungen korreliert die Testmotivation vor und nach dem Test mit .88. Zudem korreliert die Testmotivation vorher negativ (-.25) mit dem Gesamtwert des LPS-neu und die Testmotivation nachher positiv mit .46. Bei der negativen Korrelation zwischen der TMv und dem LPS-neu Gesamtwert handelt es sich um einen Suppressionseffekt. Da die TMv und TMn viel gemeinsame Varianz aufweisen, werden durch den negativen Wert irrelevante Varianzanteile der Testmotivation unterdrückt. So können durch das negative Betagewicht zu hoch geschätzte Werte etwas nach unten korrigiert werden und infolgedessen wird eine bessere Vorhersage durch die TMn möglich. Bei Betrachtung der bivariaten Korrelationen nach Bravais-Pearson für die Version A ist der Zusammenhang zwischen TMv und LPS-neu Gesamtwert zwar gering ($r = .08$; $p = .31$), aber entgegen der Modellrechnung nicht negativ. Für dieses Modell ergibt sich ein χ^2 von 195.19 mit $df = 128$ und p von .00. Für den RMSEA liegt der berechnete Wert von .056 mit 90 % Wahrscheinlichkeit im Intervall [.040; .071]; damit bleibt der Wert unter der kritischen Grenze. Der CFI liegt mit .903 knapp ober-

halb der Grenze von .90. Zusammen sprechen die Indizes für einen adäquaten Modell-Fit.

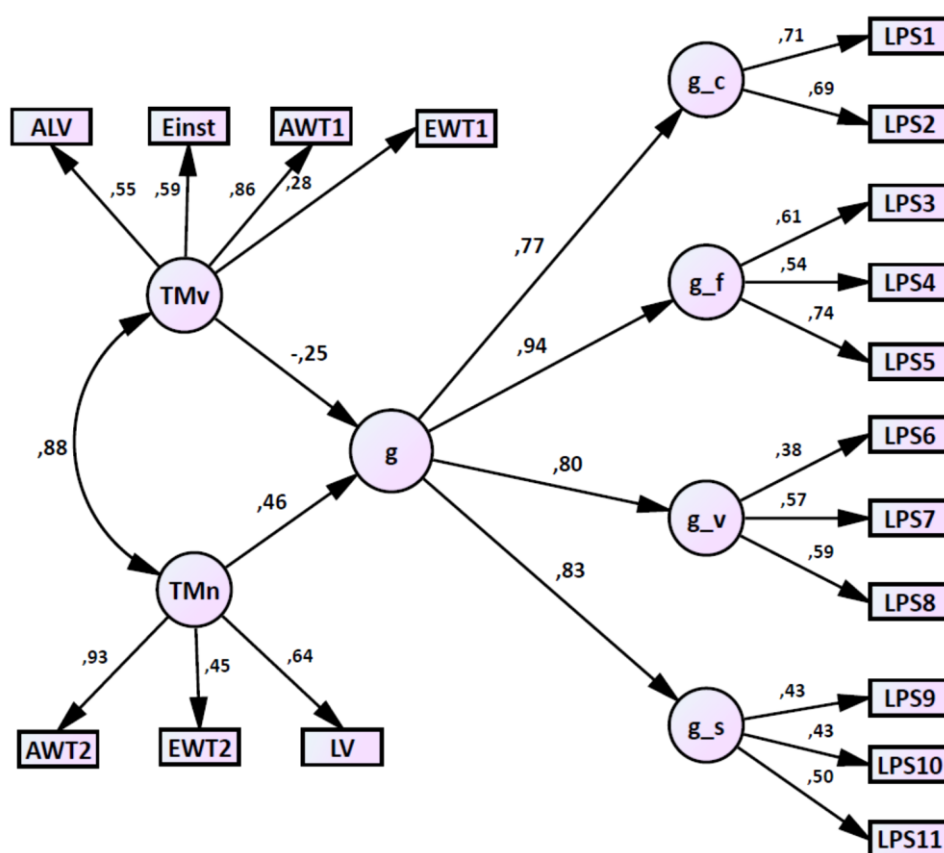


Abbildung 4.8

Modell über den Zusammenhang zwischen der Testmotivation vor und nach dem Test und deren Einfluss auf das LPS-neu (Version A)

Das in Abbildung 4.8 dargestellte Modell wurde auch für die LPS-neu Version B berechnet (siehe Anhang D-2). In diesem korreliert die Testmotivation vor und nach dem Test mit .72 und die Testmotivation vorher negativ (-.23) mit dem Gesamtwert des LPS-neu. Auch bei dieser negativen Korrelation liegt ein Suppressionseffekt vor, da bei bivariater Berechnung der Zusammenhang, wie bei Version A, zwar gering ($r = .07$; $p = .39$), aber entgegen der Modellrechnung ebenfalls nicht negativ ist. Die Testmotivation nachher korreliert hingegen positiv (.45) mit dem LPS-neu Gesamtwert. Das Modell liefert ein χ^2 von 195.27 mit $df = 128$ und p von .00. Für den RMSEA liegt der berechnete Wert von .057 mit 90 % Wahrscheinlichkeit im Intervall [.040; .072], womit er unter der kritischen Grenze bleibt. Der CFI liegt mit einem Wert von .898 knapp unter-

halb der Grenze von .90, weshalb aber trotzdem auch bei dieser Version von einem adäquaten Modell-Fit gesprochen werden kann.

Zusammenfassend lässt sich festhalten, dass sowohl die signifikant positive Korrelation als auch fast alle Modellrechnungen dafür sprechen, dass die Testmotivation einen Einfluss auf die Leistungen in einem Leistungstest hat, weshalb die Hypothese 6 an dieser Stelle als bestätigt gelten kann. Zudem hat die TMn einen höheren Einfluss auf den LPS-neu Gesamtwert als die TMv.

Im weiteren Verlauf wurde der Zusammenhang zwischen Testmotivation und dem Leistungsprüfsystem-neu, entsprechend der nachfolgenden Hypothesen 7 und 8, etwas genauer untersucht.

- H7: Der Zusammenhang zwischen Testmotivation und Intelligenz ist durch einen Leistungsanstieg aufgrund höherer Testmotivation begründet. Bei hoher Testmotivation tendiert die Korrelation gegen Null.*
- H8: Sind die Versuchspersonen für die Bearbeitung des LPS-neu ausreichend motiviert, dann ist die Intelligenz für das Testergebnis in diesem Test verantwortlich.*

Wie bereits in Kapitel 4.4.1 musste auch für diese Hypothesen die Testmotivation in Gruppen aufgeteilt werden, damit zwischen niedriger, ausreichender und hoher Testmotivation unterschieden werden konnte. Dementsprechend findet eine Einteilung in drei Gruppen statt, wobei die unteren 25 % nicht ausreichend testmotiviert, die mittleren 50 % ausreichend testmotiviert und die oberen 25 % sehr testmotiviert sind. Die Aufteilung mit Angabe von Stichprobengröße und Mittelwert kann der Tabelle 4.11 entnommen werden.

Tabelle 4.11

Einteilung der Testmotivation in die Gruppen: nicht ausreichend, ausreichend und sehr hoch testmotiviert

	TM nicht ausreichend (n = 83)	TM ausreichend (n = 166)	TM sehr hoch (n = 83)
	Mittelwert (SD)	Mittelwert (SD)	Mittelwert (SD)
TMg	133.02 (13.00)	167.17 (9.61)	198.45 (10.97)
TMv	80.18 (10.30)	97.26 (7.98)	115.99 (7.36)
TMn	52.83 (8.21)	69.91 (8.14)	82.46 (8.31)

Anmerkung. SD = Standardabweichung

Da sich die Testmotivation, vorher und nachher, aufgeteilt nach Gruppen, genau wie die gesamte Testmotivation verhält (siehe Tabelle 4.11), wurde in den folgenden Berechnungen nur die gesamte Testmotivation berücksichtigt.

Der erwartete Zusammenhang zwischen Testmotivation und Intelligenz, welcher in Hypothese 7 formuliert ist, kann der Abbildung 2.4 in Kapitel 2.4.2 entnommen werden. Dementsprechend sollte eine geringe Testmotivation auch zu geringeren Intelligenztestwerten führen. Steigt die Testmotivation in einem Test an, müsste dieses auch zu besseren Ergebnissen in diesem Test führen, bis auch eine Steigerung der Testmotivation, aufgrund kognitiver Grenzen, keine Verbesserungen der Intelligenzergebnisse mehr nach sich zieht. Um die Hypothese 7 zu überprüfen, wurden zunächst die Mittelwerte und Standardabweichungen des LPS-Gesamtwertes, aufgeteilt nach den drei Gruppen der Testmotivation, berechnet (siehe Tabelle 4.12). Beide Versionen des LPS-neu können aufgrund der in Kapitel 3.4.4 gefundenen Äquivalenz, den Gesamtwert betreffend, gemeinsam ausgewertet werden.

Tabelle 4.12

Vergleich der Testmotivationsgruppen mit dem Gesamtwert des LPS-neu

LPS_G	<i>M (SD)</i>	Min.	Max.
TM nicht ausreichend	216.04 (38.43)	102	311
TM ausreichend	229.02 (34.14)	126	331
TM sehr hoch	235.45 (32.30)	153	302

Anmerkungen. *M* = Mittelwert; *SD* = Standardabweichung

Aus Tabelle 4.12 wird ersichtlich, dass sich die drei Testmotivationsgruppen in ihren LPS_G Mittelwerten voneinander unterscheiden. So weist die Gruppe mit der geringsten Testmotivation im Mittel auch die geringsten Leistungen im LPS-neu auf, gefolgt von den ausreichend Testmotivierten. Die besten Leistungen im LPS-neu weisen die am höchsten testmotivierten Probanden auf. Somit kann an dieser Stelle festgehalten werden, dass die Leistungen im LPS-neu mit höherer Testmotivation ansteigen und der Zusammenhang zwischen Testmotivation und Intelligenz durch einen Leistungsanstieg aufgrund höherer Testmotivation begründet ist. Doch unterscheiden sich die drei Gruppen der Testmotivation in den Mittelwerten der allgemeinen Intelligenz signifikant voneinander?

Zur Beantwortung dieser Frage wurden die drei Testmotivationsgruppen einer univariaten Varianzanalyse unterzogen. Diese ergab für die gesamte Testmotivation ein signi-

fikantes Ergebnis ($F(2) = 6.82$; $p = .00$; $\eta^2 = .04$), so dass deutlich wird, dass sich die Gruppen signifikant voneinander unterscheiden. Bei Betrachtung des Scheffé Post-Hoc-Tests wird klar, welche Testmotivationsgruppen sich in der Ausprägung ihrer Leistungen im LPS-neu Gesamtwert signifikant voneinander unterscheiden (siehe Tabelle 4.13).

Tabelle 4.13

Mittelwertsdifferenzen und Signifikanzen zwischen den Testmotivationsgruppen in Abhängigkeit vom LPS-neu Gesamtwert

LPS-G		Mittlere Differenz	Standfehler	<i>p</i>
TM n.a.	TM a.	-12.99	4.68	.02
	TM s.h.	-19.41	5.41	.00
TM a.	TM n.a.	12.99	4.68	.02
	TM s.h.	-6.42	4.68	.39
TM s.h.	TM n.a.	19.41	5.41	.00
	TM a.	6.42	4.68	.39

Anmerkungen. TM n.a. = Testmotivation nicht ausreichend; TM a. = Testmotivation ausreichend; TM s.h. = Testmotivation sehr hoch; *p* = Signifikanzwert

Tabelle 4.13 zeigt, dass sich die Gruppe mit der niedrigsten Testmotivation auch signifikant von den Leistungen im Leistungstest der am höchsten Testmotivierten unterscheidet. Die Gruppe der ausreichend Testmotivierten differiert hingegen nicht signifikant von der Gruppe der Hoch-Testmotivierten.

Um zu überprüfen, ob bei hoher Testmotivation der Zusammenhang zwischen den Leistungen im LPS-neu gegen Null tendiert, wurden bivariate Korrelationen nach Barvais-Pearson mit den drei Testmotivationsgruppen und dem LPS-G berechnet. Dabei ergibt sich für die Gruppe der nicht ausreichend testmotivierten Versuchspersonen eine nicht signifikante negative Korrelation von $r = -.14$ ($p = .22$). Der Zusammenhang zwischen diesen beiden ist somit annähernd Null. Eine ähnliche Korrelation findet sich auch bei den ausreichend Testmotivierten ($r = -.06$; $p = .48$); auch hier ist der Zusammenhang zwischen dem LPS-Gesamtwert und der Testmotivation quasi Null. Eine positive, jedoch nicht signifikante Korrelation liegt bei den am höchsten testmotivierten Versuchspersonen mit $r = .15$ ($p = .18$) vor. Auch dieser Zusammenhang ist annähernd Null. Festzuhalten bleibt demnach, dass der Zusammenhang in der Gruppe der Hoch-Testmotivierten am höchsten ist, jedoch nicht signifikant wird, da auch hier die Korrela-

tion gegen Null tendiert. Zur Veranschaulichung der Zusammenhänge können die jeweiligen Regressionsgeraden der Abbildung 4.9 entnommen werden.

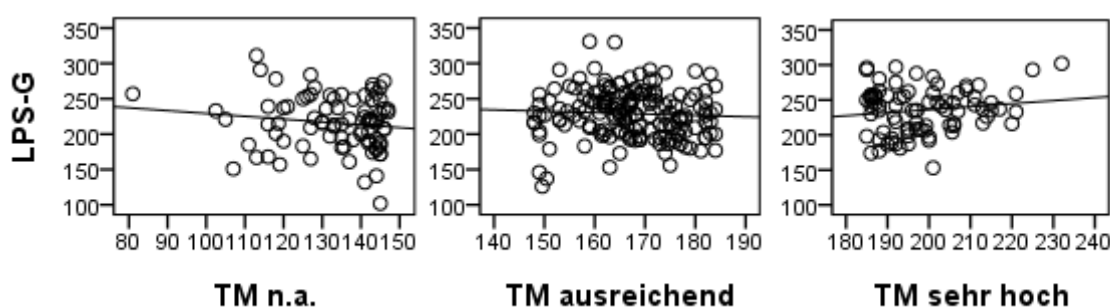


Abbildung 4.9

Regressionsgeraden über den Zusammenhang zwischen den Gruppen der Testmotivation mit dem Gesamtwert des LPS-neu

Es bleibt festzuhalten, dass sich der Zusammenhang zwischen Testmotivation und Intelligenz durch einen Leistungsanstieg aufgrund höherer Testmotivation begründen lässt. Die Hypothese 7 kann jedoch an dieser Stelle nicht bestätigt werden. Bei hoher Testmotivation tendiert die Korrelation zwar gegen Null, aber in der Tendenz führt eine höhere Testmotivation auch zu höheren Leistungen im LPS-neu Gesamtwert, was gegen die Annahme in Hypothese 7 spricht. Auch der vermutete Zusammenhang, der in Abbildung 2.4 dargestellt ist, konnte so nicht gefunden werden (siehe Abbildung 4.9).

Um die Hypothese 8 überprüfen zu können, nach der die Intelligenz für das Testergebnis verantwortlich ist, sofern die Versuchspersonen für die Bearbeitung des Leistungstests ausreichend motiviert sind, wurden die beiden Gruppen ausreichend und sehr hoch testmotiviert, da diese als insgesamt ausreichend motiviert zu bezeichnen sind, zusammengefasst und ebenfalls einer bivariaten Korrelation nach Barvais-Pearson unterzogen. Gemäß der Hypothese müsste auch die Korrelation dieser Testmotivationsgruppe mit dem LPS-Gesamtwert gegen Null tendieren. Es ergibt sich ein Zusammenhang von $r = .08$ ($p = .19$), weshalb die Hypothese 8 als bestätigt gelten kann, da die Korrelation nicht signifikant ist und somit die Intelligenz und nicht die Testmotivation für das Testergebnis im LPS-neu verantwortlich ist.

4.4.3 Schulleistung, besuchte Schulart bzw. Jahrgangsstufe und Intelligenz

Die Schulnoten in Haupt- und Nebenfächern, die besuchte Schulart und die Jahrgangsstufe wurden in der Untersuchung erhoben, um sie im Hinblick auf ihre Zusammenhän-

ge mit den Intelligenztestergebnissen untersuchen zu können. Dabei wurde zunächst die Hypothese 9 getestet.

H9: Schüler/-innen mit einer guten Schulleistung weisen höhere Werte in der allgemeinen Intelligenz des LPS-neu auf als Schüler/-innen mit einer schlechteren Schulleistung.

Die Schulleistung wurde aus dem Notendurchschnitt über alle Fächer hinweg berechnet. Sofern bei den Versuchspersonen eine Note in den Haupt- bzw. Nebenfächern fehlte, wurde diese durch Mittelwerte ersetzt. Insgesamt konnten somit für die Berechnungen des Zusammenhangs 343 Datensätze verwendet werden. Der Abbildung 4.10 kann die Verteilung der Gesamtnotendurchschnittswerte für die berücksichtigten Versuchspersonen anhand einer Häufigkeitsverteilung entnommen werden. Dabei war die beste Gesamtnote 1.09 und die schlechteste 4.11; der Mittelwert lag bei 2.71 ($SD = .55$).

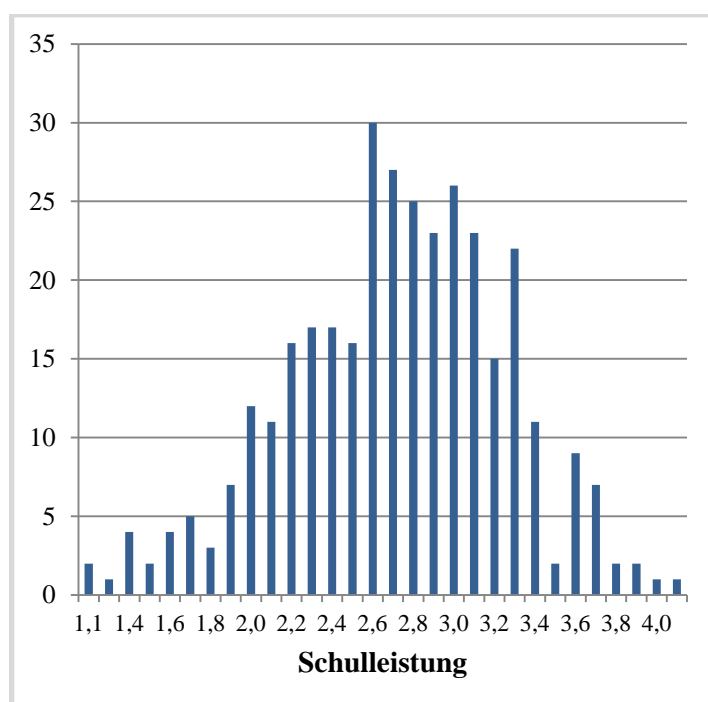


Abbildung 4.10

Häufigkeitsverteilung der Schulleistung, über alle Schulen hinweg

Die Normalverteilung des Gesamtnotendurchschnitts konnte mithilfe des Kolmogorov-Smirnov-Tests (asymptotische Signifikanz (2-seitig) = .28) bestätigt werden. Wird nun die bivariate Korrelation nach Bravais-Pearson zwischen der Schulleistung und dem

Gesamtwert des LPS-neu betrachtet, so ergibt sich zwar ein signifikanter negativer, aber eher geringer Zusammenhang ($r = -.14$; $p = .01$). Somit zeigen intelligentere Schüler/-innen auch bessere Schulleistungen, jedoch ist dieser Zusammenhang als relativ gering einzuschätzen.

Um dem Sachverhalt, dass Schulleistung und Notenvergabe zwischen den verschiedenen Schulen unterschiedlich sind, gerecht zu werden, erfolgten die weiteren Auswertungen schulartspezifisch.

In Tabelle 4.14 werden die Zusammenhänge zwischen den schulartspezifischen Schulleistungen und den Leistungen des LPS-neu etwas differenzierter betrachtet. Auch hier wurden bei den Berechnungen mit den Stratum II Komponenten des LPS-neu die Versionen A und B getrennt voneinander ausgewertet.

Tabelle 4.14

Korrelation der gemittelten Schulleistung mit dem LPS-neu

Mittelschule (n = 38)	Version A Schulleistung	Version B Schulleistung	Version A und B Schulleistung
LPS_G	-.39	-.48**	-.44**
LPS_Gc	-.25	-.47*	-
LPS_Gf	-.36	-.45	-
LPS_Gv	-.38	-.29	-
LPS_Gs	-.29	-.40	-
Mädchenrealschule (n = 85)	Schulleistung	Schulleistung	Schulleistung
LPS_G	-.07	-.29	-.17
LPS_Gc	-.22	-.11	-
LPS_Gf	-.09	-.10	-
LPS_Gv	-.17	-.24	-
LPS_Gs	.22	-.20	-
Knabenrealschule (n = 87)	Schulleistung	Schulleistung	Schulleistung
LPS_G	.12	-.23	-.07
LPS_Gc	.15	-.08	-
LPS_Gf	.09	-.10	-
LPS_Gv	.04	-.26	-
LPS_Gs	.06	-.19	-
Gymnasium (n = 133)	Schulleistung	Schulleistung	Schulleistung
LPS_G	.00	-.10	-.07
LPS_Gc	.18	.05	-
LPS_Gf	.09	-.08	-
LPS_Gv	-.14	-.08	-
LPS_Gs	-.10	-.19	-

Anmerkungen. **) Die Korrelation ist auf dem .01 Niveau signifikant; *) Die Korrelation ist auf dem .05 Niveau signifikant

Für die Berechnungen der Zusammenhänge zwischen den schulartspezifischen Schulleistungen und den einzelnen Dimensionen des LPS-neu wird lediglich die kristalline Intelligenz des LPS-neu, Version B, an der Mittelschule signifikant ($r = -.47$; $p = .05$; siehe Tabelle 4.14). Alle anderen Dimensionen weisen keine statistisch signifikanten Werte mit der schulartspezifischen Schulleistung auf. Tendenziell zeigt die Stratum II Komponente visuelle Wahrnehmung die höchsten negativen Korrelationen über alle Schularten hinweg. Darüber hinaus weist die Version B über fast alle Stratum II Komponenten des LPS-neu hinweg und auch beim LPS-Gesamtwert höhere Korrelationen mit der Schulleistung auf als die Version A. Hinsichtlich des Gesamtscores des LPS-neu (siehe Tabelle 4.14) ergibt sich ein Wert von $r = -.07$ sowohl für das Gymnasium als auch für die Knabenrealschule. Der höchste Wert zeigt sich an der Mittelschule ($r = -.44$), gefolgt von der Mädchenrealschule ($r = -.17$). Somit ist der Zusammenhang von allgemeiner Schulleistung und Intelligenz an der Mittelschule mit Abstand am höchsten, während er an den anderen Schulen nur sehr gering ist.

Aufgrund dieser sehr geringen Korrelationen zwischen der Schulleistung und den Komponenten des LPS-neu wurde darauf verzichtet, diesen Zusammenhang mithilfe der Modellberechnungen von AMOS zu überprüfen. Es bleibt festzuhalten, dass nur das Ergebnis der Mittelschule signifikante Zusammenhänge zwischen den allgemeinen Intelligenzleistungen und der Schulleistung aufweist, weshalb die Hypothese 9 abgelehnt werden muss.

H10: Die Hauptfächernoten der Versuchspersonen eignen sich als Prädiktor für die allgemeine Intelligenz besser als die Noten aus den Nebenfächern.

Zur Beantwortung der Annahmen in Hypothese 10 wurden zunächst die gemittelten Hauptfächernoten mit dem LPS-neu Gesamtwert und den Stratum II Komponenten (getrennte Berechnungen für Version A und B) nach Bravais-Pearson korreliert. Die Ergebnisse zu diesen Berechnungen zeigt Tabelle 4.15. Zudem wurden die einzelnen Hauptfächernoten und ihr Zusammenhang mit den Leistungen im revidierten Leistungsprüfsystem noch etwas differenzierter betrachtet. Diese Ergebnisse befinden sich im Anhang in Tabelle E-1 und E-2.

Tabelle 4.15

Korrelation der gemittelten Hauptfächernoten mit dem LPS-neu

	Version A	Version B	Version A und B
	g. HFN	g. HFN	g. HFN
Mittelschule (n = 38)			
LPS_G	-.50**	-.66**	-.59**
LPS_Gc	-.42	-.61**	-
LPS_Gf	-.54**	-.58*	-
LPS_Gv	-.44	-.47	-
LPS_Gs	-.24	-.53*	-
Mädchenrealschule (n = 85)	g. HFN	g. HFN	g. HFN
LPS_G	-.21	-.16	-.18
LPS_Gc	-.36*	.00	-
LPS_Gf	-.29	.06	-
LPS_Gv	-.22	-.19	-
LPS_Gs	-.15	-.16	-
Knabenrealschule (n = 87)	g. HFN	g. HFN	g. HFN
LPS_G	.08	-.20	-.07
LPS_Gc	.04	-.10	-
LPS_Gf	.04	-.11	-
LPS_Gv	.10	-.16	-
LPS_Gs	.02	-.17	-
Gymnasium (n = 133)	g. HFN	g. HFN	g. HFN
LPS_G	.04	-.12	-.06
LPS_Gc	.16	-.04	-
LPS_Gf	.14	.00	-
LPS_Gv	-.16	-.11	-
LPS_Gs	.01	-.21	-

Anmerkungen. g. HFN = gemittelte Hauptfächernoten; **) Korrelation ist auf dem Niveau von .01 signifikant; *) Korrelation ist auf dem Niveau von .05 signifikant

Aus der Tabelle 4.15 wird ersichtlich, dass der Gesamtscore des LPS-neu nur an der Mittelschule ($r = .59$; $p = .00$) signifikant negativ mit den gemittelten Hauptfächernoten korreliert. Dementsprechend weisen Schüler/-innen mit besseren Noten in den Hauptfächern an dieser Schule auch signifikant bessere Leistungen in der allgemeinen Intelligenz des LPS-neu auf. Alle anderen Schulen zeigen zwar einen negativen Zusammenhang des Hauptfächernotendurchschnitts mit dem Gesamtscore des LPS-neu, dieser ist jedoch sehr gering, wobei das Gymnasium hier den geringsten Wert ($r = -.06$; $p = .49$) zeigt. Betrachtet man nun die Zusammenhänge der Stratum II Komponenten mit den gemittelten Hauptfächernoten, werden auch diese vorwiegend nur an der Mittelschule signifikant. Den einzigen zusätzlich signifikant negativen Wert weist die Mädchenrealschule in der Version A des LPS-neu mit der kristallinen Intelligenz ($r = -.36$; $p = .02$)

auf. Alle anderen Stratum II Komponenten zeigen nur schwach negative oder leicht positive Zusammenhänge mit den gemittelten Hauptfächernoten, wobei sich die beiden LPS-neu Versionen kaum merklich voneinander unterscheiden.

Die Tabellen E-1 und E-2 (siehe Anhang) zeigen zudem die Werte für die einzelnen Hauptfächer, aus denen sich die gemittelten Hauptfächernoten berechnen. Betrachtet man hier die Zusammenhänge mit dem Gesamtscore des LPS-neu, so liefert die Version A an der Mitteschule in Deutsch den höchsten Wert ($r = -.69$; $p = .00$). An den anderen Schulen finden sich für die Version A die höchsten Korrelationen des LPS-G mit dem Fach Mathematik, wobei der Zusammenhang nur an der Mädchenrealschule ($r = -.31$; $p = .05$) signifikant wird. Werden nun die Zusammenhänge mit dem Gesamtscore des LPS-neu in der Version B betrachtet, so ergibt sich nur an der Mitteschule ein signifikantes Ergebnis im Fach Mathematik ($r = -.69$; $p = .00$). Alle anderen Schulen weisen ihre höchsten Werte mit anderen Hauptfächern auf, wobei diese jedoch nicht signifikant werden.

Im nächsten Schritt wurden die gemittelten Nebenfächernoten mit dem LPS-neu Gesamtwert und den Stratum II Komponenten nach Bravais-Pearson korreliert (siehe Tabelle 4.16).

Wie der Tabelle 4.16 zu entnehmen ist, ergeben sich für die gemittelten Nebenfächernoten fast ausschließlich geringere Werte mit dem LPS-neu Gesamtscore als für die Hauptfachnoten. Nur das Gymnasium weist ein gleiches Zusammenhangsmaß ($r = -.06$; $p = .48$) auf.

Tabelle 4.16

Korrelation der gemittelten Nebenfächernoten mit dem LPS-neu

Mittelschule (n = 38)	Version A g. NFN	Version B g. NFN	Version A und B g. NFN
LPS_G	-.26	-.26	-.26
LPS_Gc	-.10	-.28	-
LPS_Gf	-.19	-.28	-
LPS_Gv	-.28	-.11	-
LPS_Gs	-.28	-.23	-
Mädchenrealschule (n = 85)	g. NFN	g. NFN	g. NFN
LPS_G	.05	-.33*	-.13
LPS_Gc	-.09	-.18	-
LPS_Gf	.06	-.20	-
LPS_Gv	-.10	-.24	-
LPS_Gs	.24	-.20	-
Knabenrealschule (n = 87)	g. NFN	g. NFN	g. NFN
LPS_G	.14	-.23	-.05
LPS_Gc	.23	-.05	-
LPS_Gf	.13	-.07	-
LPS_Gv	-.02	-.32*	-
LPS_Gs	.09	-.18	-
Gymnasium (n = 133)	g. NFN	g. NFN	g. NFN
LPS_G	-.03	-.06	-.06
LPS_Gc	.16	.10	-
LPS_Gf	.03	-.12	-
LPS_Gv	-.10	-.05	-
LPS_Gs	-.16	-.13	-

Anmerkung. g. NFN = gemittelte Nebenfächernoten; *) Korrelation ist auf dem Niveau von .05 signifikant

Betrachtet man die Korrelationen mit den Stratum II Komponenten des LPS-neu, so liefert lediglich die visuelle Wahrnehmung für die Knabenrealschule, Version A, Signifikanz ($r = -.32$; $p = .04$) (siehe Tabelle 4.16).

Werden demnach nur die Korrelationen berücksichtigt, so sind die Hauptfächer im Vergleich zu den Nebenfächern, außer am Gymnasium, besser als Prädiktor für die allgemeine Intelligenz geeignet, abgesehen davon, dass diese außer an der Mittelschule sehr gering sind. Doch ist der Zusammenhang zwischen den Hauptfächernoten und dem LPS-neu Gesamtwert signifikant größer als der mit den Nebenfächernoten? Um diese Frage zu beantworten, wurden getrennt für die einzelnen Schulen multiple Regressionsanalysen mit den Haupt- bzw. Nebenfächernoten und dem Gesamtwert des LPS-neu (Version A und B werden zusammen betrachtet) berechnet. Welche Haupt- bzw. Ne-

benfächer der jeweiligen Schulen bei den Berechnungen berücksichtigt wurden, kann dem Kapitel 3.1.4 entnommen werden. Die entsprechende Anzahl steht in Tabelle 4.17.

Tabelle 4.17

Regressionen der Haupt- bzw. Nebenfächernoten mit dem Gesamtscore des LPS-neu

		Prädiktoren	R^2	Korr. R^2	SE	F	p
Mittelschule	HFN	3	.40	.35	31.38	7.49	.00
	NFN	4	.16	.06	37.66	1.56	.21
	HFN und NFN	7	.47	.34	31.50	3.72	.01
M.realschule	HFN	3	.10	.06	30.37	2.84	.04
	NFN	6	.07	.00	31.37	.98	.45
	HFN und NFN	9	.17	.08	30.15	1.76	.09
K.realschule	HFN	4	.02	-.03	30.26	.46	.77
	NFN	5	.04	-.02	30.23	.61	.70
	HFN und NFN	9	.06	-.06	30.70	.50	.87
Gymnasium	HFN	4	.02	-.01	32.14	.78	.54
	NFN	7	.06	.01	31.85	1.22	.30
	HFN und NFN	11	.08	.00	32.07	.97	.47

Anmerkungen. HFN = Hauptfächernoten; NFN = Nebenfächernoten; R^2 = Bestimmtheitsmaß; Korr. R^2 = korrigiertes R-Quadrat; SE = Standardfehler; F = Testwert; p = Signifikanzwert;

Betrachtet man in der Tabelle 4.17 die Regressionsanalysen aller Schulen, so findet sich das höchste Bestimmtheitsmaß (R^2), wie auch schon die Korrelationsberechnungen vermuten ließen, bei den Hauptfachnoten der Mittelschule ($R^2 = .40$). Dementsprechend können die drei Hauptfächernoten hier 40 % der Varianz des LPS-neu Gesamtscores erklären. Das Ergebnis der Regressionsanalyse wird signifikant ($p = .00$) und stellt somit den bedeutenden Zusammenhang zwischen Hauptfachnoten und dem LPS-neu Gesamtwert an der Mittelschule heraus. Die Regressionsanalyse der vier Nebenfachnoten fällt an dieser Schule hingegen schlechter aus ($R^2 = .16$; $p = .21$). Da die Hauptfächer viel Varianz erklären können und die Nebenfächer nur wenig, sind sie an der Mittelschule besser zur Vorhersage der Leistung im LPS-neu geeignet.

Auch die Ergebnisse der Mädchenrealschule deuten in dieselbe Richtung, wenngleich die erklärte Varianz der Hauptfächernoten bzw. der Nebenfächernoten am Gesamtscore des LPS-neu mit 10 % bzw. 7 % an dieser Schulen weitaus geringer ausfällt als an der Mittelschule. Dennoch wird die Regressionsanalyse der drei Hauptfächernoten auch an dieser Schule signifikant ($p = .04$), sodass hier ebenfalls die Hauptfächer als Prädiktor besser geeignet sind.

Die Ergebnisse der beiden verbleibenden Schulen deuten in eine andere Richtung. Sowohl an der Knabenrealschule als auch am Gymnasium können die Nebenfächer mit 4 % bzw. 6 % deskriptiv mehr Varianz des Gesamtscores im LPS-neu erklären als die Hauptfächernoten (2 %). Allerdings können an der Knabenrealschule bzw. am Gymnasium weder die Nebenfächernoten ($p = .70$ bzw. $p = .30$) noch die Hauptfächernoten ($p = .77$ bzw. $p = .54$) einen signifikanten Beitrag zur Vorhersage des Kriteriums leisten.

So lässt sich auch nach Durchführung der Regressionsanalysen festhalten, dass die Hauptfächernoten zumindest an zwei Schulen höhere und signifikante Zusammenhänge mit den Intelligenzleistungen aufweisen als die Nebenfächernoten. Da sich jedoch die Hauptfächernoten nicht an jeder Schule besser als Prädiktor für die allgemeine Intelligenz eignen, muss die Hypothese 10 teilweise abgelehnt werden.

Im Zusammenhang mit den einzelnen Schulen sollte anschließend geklärt werden, ob sich die Hypothese 11 bestätigen lässt.

H11: Die besuchte Schulart korreliert mit den Intelligenztestergebnissen.

Um zu überprüfen, ob die besuchte Schulart mit den Intelligenztestergebnissen zusammenhängt, wurden zunächst die Mittelwerte des LPS-neu Gesamtwertes getrennt für die einzelnen Schulen ermittelt, wobei die Versionen A und B zusammen betrachtet werden (siehe Tabelle 4.18).

Tabelle 4.18

Übersicht über die LPS-neu Gesamtwerte, getrennt nach Schulen und für die gesamte Stichprobe

	N	M (SD)	Minimum	Maximum
Mittelschule	38	201.05 (38.77)	126	276
Knabenrealschule	90	222.40 (30.19)	153	293
Mädchenrealschule	98	221.89 (33.73)	102	296
Gymnasium	135	241.84 (32.26)	151	331
Gesamt	361	227.28 (35.21)	102	331

Anmerkungen. M = Mittelwert; SD = Standardabweichung

Wie der Tabelle 4.18 zu entnehmen ist, gibt es Unterschiede zwischen den Schulen in der Höhe ihres LPS-neu Gesamtwertes, wobei sich dabei die Standardabweichungen

über die Schularten hinweg kaum merklich unterscheiden. Wie augenscheinlich diese Unterschiede zwischen den Schulen sind, wird in der Abbildung 4.10 ersichtlich.

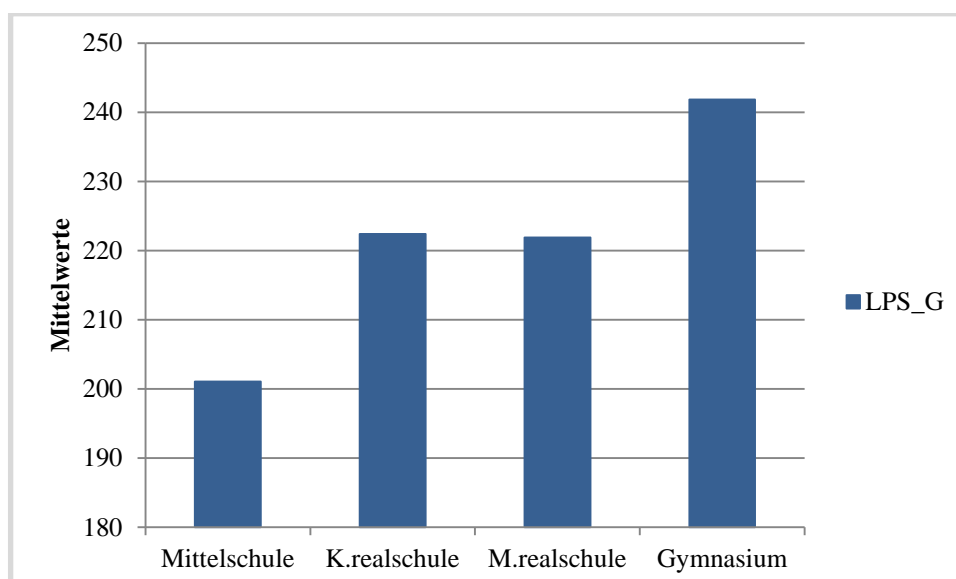


Abbildung 4.10

Übersicht über die Mittelwerte von LPS-G, getrennt nach einzelnen Schulen

Doch sind diese Unterschiede in den Mittelwerten des LPS-neu Gesamtwertes zwischen den Schularten statistisch signifikant?

Um die Mittelwerte der verschiedenen Schulen miteinander vergleichen zu können, wurde eine einfaktorielle Varianzanalyse mit der abhängigen Variable des LPS-neu Gesamtwertes und der unabhängigen Variable Schulart berechnet. Diese weist ein signifikantes Ergebnis ($F(3) = 18.39, p = .00$) auf, sodass davon auszugehen ist, dass die Mittelwerte des LPS-neu Gesamtwertes zwischen den verschiedenen Schulen signifikant voneinander abweichen. Nach Durchführung des Post-Hoc-Tests (Scheffé-Prozedur) wird klar, dass sich nur die Mittelwerte der beiden Realschulen nicht signifikant unterscheiden, die restlichen Mittelwerte hingegen schon (siehe Tabelle 4.19).

Tabelle 4.19

Mittelwertsdifferenzen und Signifikanzen zwischen den einzelnen Schulen in Abhängigkeit vom LPS-neu Gesamtwert

		<i>Mittlere Differenz</i>	<i>p</i>
Mittelschule	K.realschule	-21.35	.01
	M.realschule	-20.84	.01
	Gymnasium	-40.79	.00
K.realschule	Mittelschule	21.35	.01
	M.realschule	.51	1.00
	Gymnasium	-19.44	.00
M.realschule	Mittelschule	20.84	.01
	K.realschule	-.51	1.00
	Gymnasium	-19.95	.00
Gymnasium	Mittelschule	40.79	.00
	K.realschule	19.44	.00
	M.realschule	19.95	.00

Anmerkung. *p* = Signifikanzwert

Es kann aufgrund der Datenauswertung davon ausgegangen werden, dass die Intelligenz einen Einfluss auf die besuchte Schulart ausübt. Auf eine Modellrechnung mithilfe von AMOS, zwischen Intelligenz und deren Einfluss auf die Schulart, in der jede Schule einzeln präsentiert, das Modell jedoch über alle Schulen hinweg berechnet wird, musste aufgrund der zu geringen Stichprobengrößen der aufgeteilten Schulen an dieser Stelle verzichtet werden.

Festzuhalten bleibt demnach, dass die besuchte Schulart einen Einfluss auf die Intelligenzleistungen hat. Dabei weist das Gymnasium die höchsten Werte auf, gefolgt von den beiden Realschulen, die sich nicht voneinander unterscheiden. Die geringsten Werte des LPS-neu Gesamtscores zeigen sich an der Mittelschule. Die Hypothese 11 kann folglich als bestätigt gelten.

Des Weiteren wurde überprüft, ob die besuchte Jahrgangsstufe einen Einfluss auf die Intelligenztestergebnisse hat und ob ein Rückgang der Korrelation zwischen Schulleistung und Intelligenz mit steigendem Alter und Ausbildungsniveau, entsprechend der Jahrgangsstufe und Schulart, vorliegt.

H12: Die besuchte Jahrgangsstufe hat einen Einfluss auf die Intelligenztestergebnisse.

H13: Die Korrelation zwischen Schulleistung und Intelligenz sinkt mit steigendem Alter und Ausbildungsniveau.

Um die Hypothese 12 zu überprüfen, wurden zunächst für jede Schule einzeln die Mittelwerte des LPS-neu Gesamtwertes in Abhängigkeit von den Klassen errechnet und im Anschluss daran einfaktorielle Varianzanalysen durchgeführt. Die dabei für jede Klasse errechneten Mittelwerte sowie sich signifikant voneinander unterscheidende Klassen für jede Schule einzeln, können der Tabelle 4.20 entnommen werden.

Tabelle 4.20

LPS-G Mittelwerte der einzelnen Klassen

Mittelschule	<i>M</i>	<i>SD</i>	Minimum	Maximum
9b-R (n = 14)	179.07	31.85	126	246
9c-M (n = 11)	216.18	49.92	141	276
10-M (n = 13)	211.92	23.61	181	251
Knabenrealschule	<i>M</i>	<i>SD</i>	Minimum	Maximum
9a (n = 27)	228.37	29.21	168	283
9b (n = 21)	219.14	33.19	153	270
9c (n = 13)	222.69	27.07	180	271
9d (n = 10)	229.30	33.94	192	293
9e (n = 19)	213.68	28.49	163	256
Mädchenrealschule	<i>M</i>	<i>SD</i>	Minimum	Maximum
9a (n = 27)	226.59	29.20	161	296
9b (n = 19)	211.21	37.54	102	258
9c (n = 20)	234.35	29.34	191	291
10a (n = 7)	201.00	33.85	146	232
10b (n = 6)	243.50	24.02	217	287
10c (n = 19)	213.63	36.27	137	279
Gymnasium	<i>M</i>	<i>SD</i>	Minimum	Maximum
9a (n = 25)	237.96	25.22	187	277
9b (n = 24)	233.42	29.03	193	291
9c (n = 20)	220.75	32.81	151	274
10a (n = 27)	251.11	30.76	174	330
10b (n = 19)	258.68	33.97	177	331
10ce (n = 20)	249.35	31.77	198	311

Anmerkungen. *M* = Mittelwert; *SD* = Standardabweichung; signifikante Unterschiede zwischen Klassen einzelner Schulen sind kenntlich gemacht

Für die Mittelschule ergab die einfaktorielle Varianzanalyse ($F(2) = 4.23$, $p = .02$) ein signifikantes Ergebnis. Nach Durchführung des Post-Hoc-Tests (Scheffé-Prozedur) wird klar, dass sich die Mittelwerte von 9b-R und 9c-M signifikant voneinander unterscheiden. An der Knabenrealschule zeigte sich hingegen für die einfaktorielle Varianzanalyse ($F(4) = .85$, $p = .50$) kein signifikantes Ergebnis. Die Mittelwerte unterscheiden sich demnach nicht signifikant. Das gleiche Ergebnis ergab auch die Varianzanalyse für die Mädchenrealschule ($F(5) = 2.46$, $p = .04$). Die einfaktorielle Varianzanalyse des Gymnasiums ($F(5) = 4.27$, $p = .00$) wurde hingegen wieder signifikant. Nach Durchführung des Post-Hoc-Tests (Scheffé-Prozedur) wird deutlich, dass sich der Mittelwert der Klasse 9c signifikant von der Klasse 10b unterscheidet. Zudem lieferte der Mittelwertsvergleich zwischen Klasse 9c und 10a ein annähernd signifikantes Ergebnis.

Wie Tabelle 4.20 zu entnehmen ist, unterscheiden sich die Ergebnisse des LPS-neu Gesamtwertes nur in vereinzelten Klassen signifikant voneinander. Doch besteht auch ein Unterschied zwischen den Mittelwerten der Intelligenztestergebnisse und den beiden Jahrgangsstufen insgesamt?

Zur Überprüfung der Mittelwertsunterschiede zwischen den neunten und zehnten Klassen wurden zunächst die Mittelwerte der Jahrgangsstufen berechnet (siehe Tabelle 4.21) und im Anschluss ein t -Test für unabhängige Stichproben durchgeführt. Der Levene-Test auf Varianzhomogenität ($F = .82$; $p = .37$) wurde nicht signifikant. Es ergab sich somit für den t -Test ($t(359) = -3.75$, $p = .00$) ein signifikantes Ergebnis. Die beiden Jahrgangsstufen unterscheiden sich in ihren allgemeinen Intelligenzleistungen, wobei die zehnte Klasse im Mittel höhere Ergebnisse erzielt (siehe Tabelle 4.21).

Tabelle 4.21

Übersicht über die Mittelwerte des LPS-neu Gesamtwertes der neunten und zehnten Jahrgangsstufen

	Klasse	N	M	SD	Minimum	Maximum
LPS_G	9	250	222.74	33.51	102	296
	10	111	237.51	36.91	137	331

Anmerkungen. M = Mittelwert; SD = Standardabweichung

Aus den Daten geht hervor, dass die besuchte Jahrgangsstufe einen Einfluss auf die allgemeinen Intelligenztestleistungen hat. Um jedoch auszuschließen, dass dieser Einfluss

daher kommt, dass die älteren in den höheren Klassen auch höhere Testergebnisse haben, wurde eine univariate Varianzanalyse zur Untersuchung von Haupt- und Interaktionseffekten hinsichtlich des Alters und der Jahrgangsstufen mit der allgemeinen Intelligenz durchgeführt. Diese lieferte folgende Ergebnisse: Es ergab sich ein signifikanter Haupteffekt des Alters und der Jahrgangsstufen (siehe Tabelle 4.22). Der Interaktionseffekt zwischen diesen beiden lieferte jedoch kein signifikantes Ergebnis.

Tabelle 4.22

Ergebnisse der univariaten Varianzanalysen von Alter und Jahrgangsstufe für den Gesamtwert des LPS-neu

LPS_G (n = 361)	<i>F</i>	<i>p</i>	$p\eta^2$
Alter	2.51	.03	.03
Jahrgangsstufe	13.53	.00	.04
Alter * Jahrgangsstufe	.58	.56	.00

Anmerkungen. *F* = Testwert; *p* = Signifikanzwert; $p\eta^2$ = Effektgröße partielles Eta-Quadrat

Demnach weisen unterschiedliche Altersklassen auch unterschiedliche Intelligenztestergebnisse auf und auch die Jahrgangsstufen zeigen voneinander signifikant verschiedene Ergebnisse (siehe Abbildung 4.11).

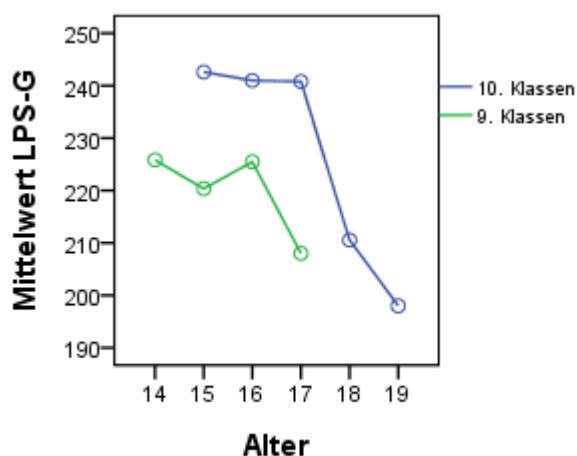


Abbildung 4.11

Univariate Varianzanalyse des LPS-neu mit Alter und Jahrgangsstufe

Auffällig ist zudem, dass die älteren Schüler/-innen entgegen dem generellen Trend des LPS-neu in dieser Studie schlechtere Ergebnisse erzielen. Um auszuschließen, dass dieses Ergebnis durch die Klassenwiederholer bedingt ist, wurde anschließend eine univa-

riate Varianzanalyse von Alter und Jahrgangsstufe für den Gesamtwert des LPS-neu ohne die Klassenwiederholer berechnet. Hierbei ergab sich ein signifikanter Haupteffekt für die Jahrgangsstufe ($F(\text{Jahrgangsstufe}) = 16.30$; $p = .00$; $p\eta^2 = .06$). Der Haupteffekt des Alters und der Interaktionseffekt zwischen diesen lieferte kein signifikantes Ergebnis ($F(\text{Alter}) = .19$; $p = .90$; $p\eta^2 = .00$; $F(\text{Alter} * \text{Jahrgangsstufe}) = .46$; $p = .50$; $p\eta^2 = .00$). Betrachtet man zudem die Abbildung 4.12, so erzielen hier die älteren Schüler/-innen keine schlechteren Ergebnisse im LPS-neu Gesamtwert.

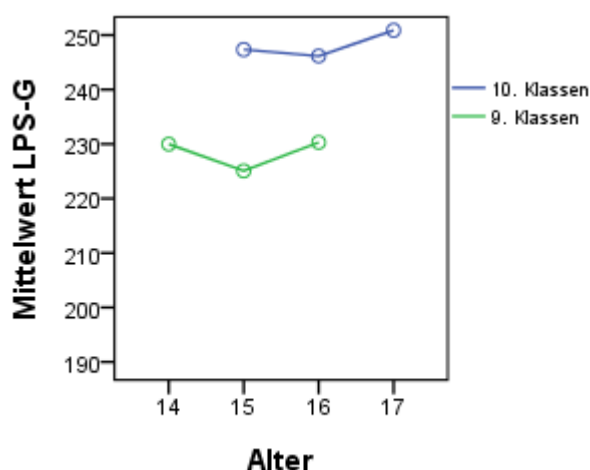


Abbildung 4.12

Univariate Varianzanalyse des LPS-neu mit Alter und Jahrgangsstufe ohne Klassenwiederholer

Das schlechtere Abschneiden der älteren Schüler/-innen im LPS-neu kann demnach durch die Klassenwiederholer erklärt werden.

Zusammenfassend lässt sich festhalten, dass die Jahrgangsstufe einen Einfluss auf die allgemeine Intelligenz besitzt. Die Hypothese 12 kann folglich als bestätigt gelten.

Um auch die letzte Hypothese 13 zu überprüfen – die Korrelation zwischen Schulleistung und Intelligenz sinkt mit steigendem Alter und Ausbildungsniveau – wurden zunächst Korrelationen nach Bravais-Pearson, abhängig von den Altersklassen bzw. Jahrgangsstufen, zwischen Schulleistungen und dem LPS-neu Gesamtwert berechnet. Um auch hier dem Sachverhalt, dass die Notenvergabe an den einzelnen Schulen unterschiedlich ist, gerecht zu werden, wurden die Zusammenhänge für jede einzelne Schule getrennt berechnet.

Tabelle 4.23

Korrelation zwischen dem LPS-neu Gesamtwert und der Schulleistung in Abhängigkeit von den Altersklassen bzw. Jahrgangsstufen

LPS_G und Schulleistung	Mittelschule	M.realschule	K.realschule	Gymnasium
14 Jahre	-.68 (6)	-.45 (16)	-.03 (19)	.11 (22)
15 Jahre	-.49* (20)	-.00 (24)	-.21 (47)	-.22 (62)
16 Jahre	-.12 (10)	-.13 (34)	.44 (17)	.15 (40)
17 Jahre	-1.00** (2)	-.37 (5)	-.95 (4)	-.49 (7)
18 Jahre	-	.02 (6)	-	-
9. Jahrgangsstufe	-.56** (25)	-.22 (65)	-.07 (87)	-.15 (68)
10. Jahrgangsstufe	.17 (13)	.01 (20)	-	.09 (65)

Anmerkung. **) Die Korrelation ist auf dem Niveau von .01 signifikant; *) Die Korrelation ist auf dem Niveau von .05 signifikant; in Klammern stehen die jeweiligen Stichprobengrößen

Aus Tabelle 4.23 wird ersichtlich, dass die Korrelation zwischen dem LPS-neu Gesamtwert und der Schulleistung nicht unbedingt mit steigendem Alter sinkt. Betrachtet man hingegen die Jahrgangsstufe 10 als höheres Ausbildungsniveau im Vergleich zur 9. Jahrgangsstufe, so kann hier ein Absinken des Zusammenhangs zwischen Schulleistung und Intelligenztestleistungen festgestellt werden. Jedoch wird dieser Unterschied nur am Gymnasium, nach Berechnungen von univariaten Varianzanalysen, signifikant (siehe Tabelle 4.24). Die Knabenrealschule konnte nicht berücksichtigt werden, da hier lediglich an neunten Klassen Testungen stattfanden.

Tabelle 4.24

Ergebnisse der univariaten Varianzanalysen von Schulleistung und Jahrgangsstufe für den Gesamtwert des LPS-neu

LPS_G	F-1	df	p	$p\eta^2$	F-2	df	p	$p\eta^2$	F-3	df	p	$p\eta^2$
Schulleistung	1.51	38	.06	.44	1.05	23	.43	.32	1.62	14	.18	.59
Jahrgangsstufe	13.02	1	.00	.15	1.19	1	.28	.02	.65	1	.43	.04
Schullei. * Jahrgangs.	1.87	19	.03	.33	1.19	9	.32	.17	.51	6	.79	.16

Anmerkungen. F = Testwert; F-1 = Gymnasium; F-2 = Mädchenrealschule; F-3 = Mittelschule; df = Freiheitsgrade; p = Signifikanzwert; $p\eta^2$ = Effektgröße partielles Eta-Quadrat

Aus beiden Berechnungen wird klar, dass die Hypothese 13 nicht bestätigt werden kann. Die Korrelation zwischen Schulleistung und Intelligenz sinkt nicht signifikant mit steigendem Alter und Ausbildungsniveau.

5. Diskussion

Im Folgenden werden zunächst die verwendeten Messinstrumente kritisch erörtert und eventuell Verbesserungsvorschläge gegeben. Im Anschluss daran werden die Ergebnisse der Studie diskutiert und für mögliche zukünftige Forschungsfragen Schlussfolgerungen gezogen. Auch der Nutzen der Arbeit soll zusammenfassend dargestellt werden.

5.1 Diskussion der Methode

Kritisch diskutiert werden in diesem Kapitel die Messinstrumente, die Stichprobe sowie der Untersuchungsverlauf, da diese Aspekte bei der Interpretation der Ergebnisse eine wichtige Rolle spielen können.

5.1.1 Untersuchungsmethode

5.1.1.1 *Das Regensburger Leistungs-Motiv-Inventar für Kinder und Jugendliche (RLMI -K/J)*

Es zeigen sich für die Gesamtskalen ausreichende bis hohe Werte der internen Konsistenzen, wobei der Bereich Ausbildung etwas höhere Werte aufweist als der Bereich Freizeit. Auch die anderen Werte sprechen für eine Normalverteilung der erhobenen Daten. Aus diesen Gründen kann von einem formal geeigneten Testverfahren zur Messung des Leistungsmotivs gesprochen werden. Betrachtet man jedoch die Korrelationen untereinander (siehe Kapitel 3.4.3.2) und auch das euklidische Distanzmodell (siehe Kapitel 4.1, Abbildung 4.1), stellt sich die Frage, ob die Unterscheidung der Leistungsmotivation des RLMI-K/J in vier verschiedene Motive erforderlich ist. So verhält sich das Motiv FM sehr ähnlich zu den beiden Motiven HM und FE. Zudem weisen auch die beiden Motive HM und FE eine sehr hohe Korrelation beziehungsweise eine sehr geringe Distanz zueinander auf. Zumindest in dieser Studie musste das Motiv FM aus den weiteren Untersuchungen ausgeschlossen werden, da eine Unterscheidung zu den beiden Motiven HM und FE kaum möglich war.

Die Schüler/-innen fanden den Fragebogen an manchen Stellen schwierig, da sie sich in einige Beispielsituationen und dazugehörige Aussagen nur schwer hineinversetzen und aufgrund dessen ihre Meinung nicht äußern konnten. Hinzu kam, dass vereinzelte Schüler/-innen ihre Kreuze zu den Aussagen wahllos setzten, eventuell aus mangelnder

Motivation beziehungsweise Desinteresse am Fragebogen, nur um schneller fertig zu werden. Das könnte die Ergebnisse zusätzlich verzerrt haben.

5.1.1.2 Das Leistungsmotivgitter für Kinder und Jugendliche – Kurzversion (AMG-S K-J)

Auch dieses Messinstrument liefert gute bis sehr gute Werte der internen Konsistenzen für die einzelnen Skalen. Zudem weisen auch die zusätzlich berechneten Werte auf eine normale Verteilung der erhobenen Daten hin. Nach Betrachtung der Korrelationen (siehe Kapitel 3.4.4.2) und der euklidischen Distanzen (siehe Kapitel 4.1, Abbildung 4.1) scheint sich auch in dieser Studie das Modell der Leistungsmotivation mit drei Komponenten – HE, FMa und FMp – zu bestätigen. Untersucht wurde jedoch nicht, da der Schwerpunkt auf der Erfassung der Testmotivation und deren Zusammenhänge lag, ob Schüler/-innen mit einer hohen FMp auch schlechtere Leistungsergebnisse liefern und zudem schlechtere Schulleistungen aufweisen als Schüler/-innen mit einer hohen FMa, wie Schmalt in einigen Untersuchungen finden konnte (1976; 1999; 2005).

Das Feedback der Schüler/-innen zum Fragebogen zeigte, dass für einige die Schwierigkeit der Bearbeitung darin lag, die nicht ganz eindeutigen Bilder richtig zu interpretieren. So hätten sie eindeutige Bilder bevorzugt. Außerdem schienen einige Schüler/-innen die Instruktion nicht richtig gelesen beziehungsweise nicht richtig verstanden zu haben, da sie nur eine zutreffende Antwort ankreuzten, obwohl mehrere Kreuze möglich waren. Auch bei diesem Fragebogen gab es vereinzelt Schüler/-innen, die ihre Kreuze wahllos setzten, ohne die Aussagen richtig zu lesen.

5.1.1.3 Das neue Leistungsprüfsystem (LPS-neu)

Das revidierte Leistungsprüfsystem befindet sich noch in der Validierungsphase, wobei die in dieser Studie gewonnenen Daten ebenfalls zur Validierung verwendet werden können. Auch dieses Testverfahren liefert teststatistisch sehr gute Werte. So liegen nicht nur die internen Konsistenzen der beiden LPS-neu Versionen auf einem hohen Niveau, sondern auch die nach der Spearman-Brown-Formel korrigierten Halbierungsreliabilitäten für die einzelnen Subtests liefern sehr zufriedenstellende Ergebnisse (siehe Kapitel 3.4.6.2). Bei den einzelnen Subtests kann außerdem davon ausgegangen werden, dass sie sich entsprechend einer Normalverteilung verhalten. Da eine Normalverteilung nicht nur bei den Subtests, sondern auch beim Gesamtwert des LPS-neu vor-

liegt, erfüllt das revidierte Leistungsprüfsystem eine wichtige Eigenschaft eines Intelligenztests. Darüber hinaus ermöglichte die Anwendung des LPS-neu einen Überblick über die Verteilung der kognitiven Leistungsfähigkeiten getrennt nach Geschlecht, Alter, Schulart und Jahrgangsstufe. Die bei der Durchführung verwendeten Parallelversionen A und B verhielten sich – zumindest den Gesamtscore betreffend – äquivalent.

Zum Verhalten der Schüler/-innen während der Testung lässt sich festhalten, dass der überwiegende Teil gestresst wirkte; beispielsweise wurden Fingernägel gekaut oder mit den Füßen gewippt. So kann darauf geschlossen werden, dass die meisten der getesteten Schüler/-innen motiviert waren, sie den Test ernst nahmen und ihnen ein gutes Testresultat wichtig war. Einige waren sogar so motiviert, dass sie auch bei Ablauf der Testzeit erst nach nochmaliger Ermahnung die Bearbeitung der Subtests beendeten. Dennoch gab es auch Schüler/-innen, die weniger Interesse am Test zeigten; sie beendeten frühzeitig die Bearbeitung der Subtests oder setzten ihre Antworten wahllos.

Neben der Beobachtung des Verhaltens halfen auch die Rückmeldungen einiger Schüler/-innen, mehr über deren Gefühle während des Tests zu erfahren. So war es für einige Schüler/-innen sehr belastend, dass sie die einzelnen Aufgaben nicht zu Ende bearbeiten konnten, auch wenn in der Anfangsinstruktion darauf verwiesen wurde. Darüber hinaus hat nur ein geringer Teil der Schüler/-innen die Chance ergriffen, bei den Aufgaben, bei denen sie die Antwort nicht wussten, zu raten. Hier ließen sie die Aufgaben lieber unbearbeitet.

Die Instruktionen des LPS-neu scheinen bei einigen Subtests zu lang zu sein, sodass die Schüler/-innen teilweise nur gelangweilt zuhörten. Ihrer Meinung nach wird vor allem zu oft erklärt, was bei einem Fehler zu tun sei. Außerdem sorgten die Instruktionen des Subtests 9 und 10 in einigen Klassen für Verwirrung. So reichte einigen Schüler/-innen das Vorlesen der Instruktion für das Verstehen der Aufgabenstellung nicht aus. Sie konnten erst nach Rücksprache beziehungsweise Erläuterung der Aufgaben durch den Versuchsleiter während der Testzeit mit der eigentlichen Bearbeitung beginnen. Nach dem Vorlesen der Instruktionen bestand zwar die Möglichkeit, Fragen zu stellen, doch dies wurde trotz eventuell vorhandener Missverständnisse von den meisten Schüler/-innen nicht in Anspruch genommen.

Die Dauer der Testung wurde von vielen Schüler/-innen ebenfalls als zu lang empfunden. So konnten einige ihre Konzentration nicht bis zum Schluss aufrechterhalten und brachen die letzte Aufgabe, fast ohne eine Antwort zu geben, vorzeitig ab. Dennoch war

der überwiegende Teil der Schüler/-innen bis zum Schluss der Testung motiviert und bearbeitete den Test konzentriert.

Auch der Umstand, dass bis zu 26 Schüler/-innen gleichzeitig getestet wurden, sollte nicht unberücksichtigt bleiben. Dieses hatte zur Folge, dass auf den Einzelnen wenig Rücksicht genommen werden konnte und der Lärmpegel automatisch höher war als bei einer Einzeltestung; inwiefern das Auswirkungen auf die Leistungen im LPS-neu hatte, kann an dieser Stelle nicht beantwortet werden.

5.1.1.4 Fragebogen zur Erfassung der Testmotivation

Für diese Studie wurde ein Fragebogen zur Erfassung der Testmotivation entwickelt, da es bislang nur vereinzelte Fragebögen gibt, die dieses Konstrukt erfassen. Auch hier finden sich teilweise gute teststatistische Werte, die für die Erhebung der einzelnen Skalen der Testmotivation nach dem Modell von Eccles et al. (1983) sprechen. Dabei scheinen vor allem die beiden Skalen Aufgabenwert und Erfolgserwartung die Testmotivation der Schüler/-innen gut erfassen zu können. Die Skala Einstellungen allein lieferte hingegen keine guten Werte bei der Testanalyse. So scheint es, dass die Testmotivation nicht durch die Einstellung beziehungsweise die wahrgenommene Einstellung der Bezugspersonen beeinflusst wird, wie nach Pohlmann et al (2005) zu erwarten gewesen wäre. Es könnte aber auch sein, dass gerade bezüglich der persönlichen Einstellung Schüler/-innen mit einem Durchschnittsalter von 15 Jahren Schwierigkeiten haben, diese entsprechend ihrer tatsächlichen Meinung zu äußern (siehe hierzu Kapitel 5.1.2).

Darüber hinaus scheint es so zu sein, dass die Schüler/-innen ihre Testmotivation nach dem Test besser einschätzen können als vor dem Test. Zumindest waren hier die Zusammenhänge mit den Leistungen im LPS-neu höher (siehe hierzu Kapitel 5.2.4.2). Trotzdem ist die Erhebung der Testmotivation sowohl vor als auch nach einem Test wichtig, um auch den Einfluss der wahrgenommenen Leistung in einem Test auf die Motivation untersuchen zu können (Sanchez et al., 2000).

Festzuhalten bleibt auch, dass hinsichtlich der Strukturmodellierung zwar die meisten Parameter des Modells der Testmotivation für einen guten Modell-Fit sprechen, die berechneten Indizes hingegen nicht. Dabei stellt sich die Frage, ob das Modell im Rahmen einer neuen Strukturgleichungsberechnung überarbeitet werden sollte, wobei eventuell nicht trennscharfe Items sowie die Skala Einstellungen aus dem Modell entfernt werden könnten. Darüber hinaus kann es jedoch auch möglich sein, dass im Rahmen

des Modells der Testmotivation keine Linearität der angenommenen Zusammenhänge besteht und dieses somit letztendlich nicht für eine Strukturgleichungsberechnung geeignet ist.

Insgesamt scheint es, dass die beiden Fragebögen von den Schüler/-innen gut angenommen und gewissenhaft ausgefüllt wurden, was unter Umständen auch an der geringen Bearbeitungszeit der Bögen liegen kann. Bei einer Testung von gleichzeitig bis zu 26 Schüler/-innen sollte jedoch auf eine umweltfreundliche beidseitige Bedruckung des Fragebogens auf DIN A5 verzichtet werden. Nur so kann die vollständige Bearbeitung des Fragebogens gewährleistet und das unnötige Fehlen von Daten verhindert werden.

5.1.1.5 Fragebogen zu den soziodemographischen Daten

Dieser Fragebogen schien am Gymnasium besondere Schwierigkeiten zu bereiten, da dieser hier erst am eigentlichen Testtag ausgefüllt wurde. So wussten viele Schüler/-innen ihre Schulnoten aus dem vorangegangenen Schuljahr nicht mehr. Diese waren jedoch aufgrund der Vergleichbarkeit mit den anderen Schulen wichtig. Deshalb gaben einige Schüler/-innen nur Noten an, an die sie sich noch sicher erinnerten.

Darüber hinaus wollten einige Schüler/-innen keine Schulnoten angeben, aus Angst beziehungsweise Scham, die Anonymität der Untersuchung könnte nicht gewährleistet sein.

In weiteren Untersuchungen sollte auf die vollständige Bearbeitung eines solchen Fragebogens verstärkt geachtet werden, da viele Schüler/-innen zum Beispiel nicht bemerkten, dass sie eine Angabe, wie die der Schulart im Vorjahr oder die der Klasse im letzten Schuljahr, unausgefüllt ließen. Wünschenswert wäre es außerdem, diese beiden Punkte in einer erneuten Untersuchung zu ähnlichen Themen auch an der Grund- und Mittelschule beziehungsweise an der Hauptschule zu erfassen. Zwar scheint die Schulart und die Klasse im letzten Schuljahr an dieser Schulart nicht so relevant zu sein, könnte aber dennoch Aspekte enthalten, die aufgrund fehlender Erhebung in dieser Studie nicht berücksichtigt werden konnten.

5.1.2 Stichprobe

Hinsichtlich der Geschlechter war die Stichprobe relativ gleichmäßig verteilt, bis auf die Überzahl der männlichen Probanden am Gymnasium (99 Schüler zu 36 Schülerinnen). Betrachtet man die Teilnehmerzahl der Versuchspersonen der beiden Jahrgangs-

stufen, so wurden eindeutig mehr Schüler/-innen der neunten als der zehnten Jahrgangsstufe getestet (250 zu 111). Dieser Umstand lag zum einen an den größeren Klassen in der neunten Jahrgangsstufe. Zum anderen könnte die geringere Teilnahme der zehnten Klassen auch daran liegen, dass sich Schüler/-innen dieser Jahrgangsstufe mehr vom Verhalten ihrer Altersgruppe beeinflussen lassen, da diese eine wichtige sozialisierende Rolle einnehmen (Limbourg, 1998). Das war zum Teil sehr deutlich an der Mädchenrealschule zu beobachten; sofern sich ein Teil der Schülerinnen entschloss, nicht an der Testung teilzunehmen, taten dies auch weitere, die zunächst unentschlossen waren. Dieses Verhalten konnte hingegen in den neunten Jahrgangsstufen nicht festgestellt werden. Aus diesem Grund wäre es bei nachfolgenden Untersuchungen zu empfehlen, vor allem die Schüler/-innen der zehnten Jahrgangsstufe besonders zu motivieren, um dadurch dem Gruppendruck beziehungsweise dem Unwillen an der Testung entgegenzuwirken.

Betrachtet man zudem die untersuchte Altersstufe, so sind die Schüler/-innen der neunten Jahrgangsstufe im Durchschnitt ein Jahr jünger als die der zehnten ($M = 15.02$; ($SD = .78$); $M = 16.07$; ($SD = 1.00$). Ein Jahr kann aber gerade in der Adoleszenz, in der das Leben der Jugendlichen von der Identitätsentwicklung und -findung geprägt ist (Berk et al., 2011), sehr bedeutsam sein. Hinzu kommt, dass in der vorliegenden Untersuchung die jüngste Versuchsperson 14 und die älteste 19 Jahre alt war. Dabei kann ein mitunter 19- oder 18-jähriger Proband eine ganz andere Meinung beziehungsweise Einstellung vertreten, als er es noch mit beispielsweise 14 Jahren tat. Folglich wäre es bei nachfolgenden Untersuchungen interessant zu erfahren, ob vor allem die Skala Einstellungen des Testmotivationsfragebogens bei einer anderen Altersgruppe, zum Beispiel bei Erwachsenen, bessere teststatistische Werte erzielt hätte. Denn gerade in der Pubertät, in der eines der wichtigsten Ziele die Autonomie – eine von den Eltern getrennte Selbstdefinition der Jugendlichen (Gerhard, 2005) – ist, sind die Einstellungen der Bezugspersonen, vor allem der Eltern, nicht so wichtig. Das könnte zu den wenig trennscharfen Ergebnissen der Skala Einstellungen des Fragebogens zur Testmotivation geführt haben.

Darüber hinaus könnten die Ergebnisse von lediglich 38 Versuchspersonen der Mittelschule im Vergleich zu beispielsweise 135 Gymnasiasten/-innen durch den geringeren Stichprobenanteil verzerrt sein. Die Stichprobenverteilung der beiden Realschulen mit jeweils 90 Schülern und 98 Schülerinnen scheint im Vergleich zur Stichprobe des Gymnasiums weniger kritisch zu sein.

Das augenscheinliche Interesse an der Studie vor allem am Leistungstest machte sich sowohl in der Höhe der erfassten Testmotivation bemerkbar als auch durch die vereinzelten Nachfragen der Schüler/-innen, wann das Ergebnis abrufbar sei und ob die Lösungen für einzelne Aufgaben zur Verfügung gestellt würden. Auch zur Leistungsmotivation, den bearbeiteten Fragebögen sowie zur Entwicklung des Fragebogens zur Testmotivation wurden interessiert Fragen gestellt.

5.1.3 Untersuchungsablauf

Zum Untersuchungsablauf lassen sich zwar keine nachteiligen Unterschiede in den Ergebnissen zwischen der Mittelschule und den restlichen Schulen feststellen; dennoch wäre ein gleicher Ablauf an allen Schulen wünschenswert, dieser war jedoch nach Absprache mit den Schulleitern nicht möglich. So scheint sich das Vorgehen an der Mittelschule anzubieten: Hier wurden die Fragebögen zum Leistungsmotiv sowie zu den soziodemographischen Daten bereits vor dem Testtag bearbeitet und nur die Testung des LPS-neu sowie die Bearbeitung der Testmotivationsfragebögen fanden am eigentlichen Testtag statt. Die Fragebögen konnten so ohne Zeitdruck bearbeitet werden, für eventuell auftretende Fragen blieb mehr Zeit und auch auf Rückfragen bezüglich dieser Arbeit beziehungsweise der einzelnen Themeninhalte konnte in Ruhe eingegangen werden.

Trotz methodischer Sorgfalt bei der Untersuchungsdurchführung ließen sich nicht alle Probleme verhindern. So waren bei einer nicht geringen Anzahl von Schüler/-innen Fehler im Versuchspersonencode, der sich aus den ersten beiden Buchstaben des Vornamens des Vaters, aus den ersten beiden Buchstaben des Vornamens der Mutter und aus der Monatszahl des Geburtstages der Probanden zusammensetzte, festzustellen. Die Zuordnung der einzelnen Fragebögen und Tests zu einer Versuchsperson waren dadurch oft erschwert. Gründe für die fehlerhafte Bildung könnten unter anderem Desinteresse an der Untersuchung oder ungenaues Lesen der Instruktionen gewesen sein. In zukünftigen Untersuchungen empfiehlt es sich, auf die exakte Bildung des Versuchspersonencodes zu achten und diese auf Übereinstimmung bei allen Fragebögen von den Probanden am Schluss der Untersuchung nochmals überprüfen zu lassen.

Darüber hinaus schien in den meisten Klassen die 90-minütige Untersuchungszeit nicht auszureichen. Dadurch wurden einige Schüler/-innen, die in Ruhe die letzten Fragebögen ausfüllen wollten, unter Zeitdruck gesetzt oder beendeten diese frühzeitig, bevor sie mit der Bearbeitung fertig waren. Auch auf den Testzeitpunkt sollte in zukünfti-

gen Untersuchungen, wenn möglich, geachtet werden. So schienen Schüler/-innen, die in der dritten und vierten Unterrichtsstunde getestet wurden, am konzentriertesten zu arbeiten. Um die Testung des LPS-neu hier nicht unterbrechen zu müssen, wurde die große Schulpause auf das Ende der Untersuchung verlegt. Im Vergleich dazu schienen Schüler/-innen in den ersten und letzten beiden Schulstunden eher müde und unkonzentriert zu sein. Auch der Testtag könnte einen Einfluss auf die Motivation in einer solchen Untersuchung haben. So könnten Schüler/-innen am Anfang beziehungsweise in der Mitte der Woche motivierter sein als beispielsweise am Freitag, wo sich der überwiegende Teil schon auf das bevorstehende Wochenende freut. Ob der Testzeitpunkt und der Testtag einen Einfluss auf die Untersuchungsergebnisse haben, wurde jedoch in dieser Studie nicht untersucht.

Weiter wäre es empfehlenswert, wenn zumindest am Anfang der einzelnen Testung ein Lehrer anwesend ist. Dieser kann schneller für Ruhe in der Klasse sorgen. Denn gerade bei Gruppentestungen ist Ruhe eine Grundvoraussetzung dafür, dass sich alle Teilnehmer/-innen konzentrieren können. Ob die Untersuchungsergebnisse vor allem im LPS-neu anders ausgefallen wären, wenn die Testungen in kleineren Gruppen stattgefunden hätten, kann an dieser Stelle nicht beantwortet werden. In den großen Gruppen fiel besonders auf, dass vorwiegend Schüler/-innen in den 9. Klassen bei Schwierigkeiten mit den Aufgabenstellungen nicht nachfragten. Testungen, die in kleineren Gruppen durchgeführt werden, könnten dazu beitragen, dass Schüler/-innen dieser Altersgruppe bei Nichtverstehen früher nachfragen. Außerdem kann der Versuchsleiter bei weniger Versuchspersonen schneller erkennen, ob jemand die Aufgabenstellung nicht versteht und zeitiger helfen. Dadurch kann früher beziehungsweise überhaupt mit der Bearbeitung begonnen werden, was unter Umständen zu besseren Ergebnissen im jeweiligen Test führen kann.

5.2 Diskussion der Ergebnisse

In den folgenden Kapiteln werden die Ergebnisse dieser Studie interpretiert und mit den in Kapitel 1 vorgestellten Theorien in Verbindung gebracht sowie kritisch diskutiert. Zunächst werden die Ergebnisse der einzelnen Konstrukte und im Anschluss daran die zusammenhängenden Ergebnisse erörtert. Darüber hinaus werden mögliche zukünftige Fragestellungen abgeleitet.

5.2.1 Leistungsmotiv

Die Einbindung der Leistungsmotivkomponenten des RLMI-K/J und des AMG-S K-J in das quadripolare Modell der Leistungsmotivation von Covington et al. (Covington & Omelich, 1991; Covington & Roberts, 1994) war, bis auf das Leistungsmotiv FM aus dem RLMI-KJ, erfolgreich. Festzuhalten bleibt, dass die verbleibenden Leistungsmotive der beiden Fragebögen teilweise signifikant miteinander korrelieren (zum Beispiel HE aus beiden Bögen mit $r = .28$; $p = .00$). Es kann demnach davon ausgegangen werden, dass die in den beiden Bögen zugrunde gelegten vergleichbaren Leistungsmotive auch vergleichbare Motive erfassen. Darüber hinaus spricht auch die MDS für die Einteilung der beteiligten Motive in die vier Leistungsmotivtypen von Covington et al. (Covington & Omelich, 1991; Covington & Roberts, 1994) und liefert außerdem für das Modell mit zwei unabhängigen Dimensionen höchst zufriedenstellende Werte (Stress-Wert = .00; $R^2 = .99$). Auch konnten keine Unterschiede hinsichtlich des Alters noch des Geschlechts in Bezug auf die Leistungsmotivation beziehungsweise für die Verteilung der einzelnen Leistungsmotivtypen gefunden werden. Überdies war die Zuteilung der Versuchspersonen zu den einzelnen Clustern relativ gleichmäßig über die gesamte Stichprobe verteilt.

Zum einen lässt sich aus den gefundenen Werten und Ergebnissen schlussfolgern, dass unter gewissen Umständen die Einteilung des RLMI-K/J in vier verschiedene Leistungsmotive aufgrund zu hoher Ähnlichkeit zwischen den drei Motiven FM, HM und FE nochmals überdacht werden sollte. Zum anderen scheint es, dass durch die Unterteilung der Leistungsmotivation des quadripolaren Modells in vier verschiedene Leistungsmotivtypen eine bessere Unterscheidbarkeit und eine differenziertere Betrachtung der zugrundeliegenden Leistungsmotive möglich sind. Zwar ist auch in diesem Modell die Leistungsmotivation durch die beiden Motive FM und HE gekennzeichnet, wie es bereits beispielsweise Atkinson (1957) für sein Risikowahlmodell vorgeschlagen hat, doch kann sich durch die Behandlung der Leistungsmotivation als zwei unabhängige Dimensionen der Informationswert erhöhen (Brunstein & Heckhausen, 2006). So ermöglicht diese Einteilung auch Aussagen über die beiden Leistungsmotivtypen – des Übermotivierten und des Misserfolgsakzeptierenden –, über die sonst keine Aussagen getroffen werden könnten. Zudem ist eine differenzierte Betrachtung aufgrund der Annahme möglich, dass die Motive FM und HE bei einer Person in jeder möglichen Ausprägung auftreten können.

5.2.2 Testmotivation

Aufgrund der Modellrechnungen mit AMOS konnte die Testmotivation mit den verschiedenen Subskalen der Theorie von Eccles et al. (1983) nicht adäquat erfasst werden. So konnte sowohl für das Modell der Testmotivation mit den Skalen als beobachtbare Variablen ($\chi^2 = 102.62$ ($df = 11$; $p = .00$); RMSEA = .159; CFI = .87) als auch für die einzelnen Items aus dem Fragebogen als beobachtbare Variablen ($\chi^2 = 3234.6$ ($df = 1.315$; $p = k. A.$); RMSEA = .073; CFI = .65) von keinem adäquaten Modell-Fit ausgegangen werden. Darüber hinaus konnten auch für die getrennten Berechnungen der Testmotivation sowohl vor (TMv) als auch nach dem Test (TMn), jeweils mit Skalen beziehungsweise Items als beobachtbare Variablen, zwar bessere, aber dennoch keinesfalls angemessene Modell-Fits gefunden werden. Auch das Modell über den Zusammenhang zwischen der Testmotivation vor und nach dem Test musste als inadäquat zurückgewiesen werden.

Dessen ungeachtet bleibt festzuhalten, dass bei Betrachtung der einzelnen Skalen der Testmotivation, da diese die Summe der Skalen ist, alle auf dem .01 Niveau signifikant mit der gesamten Testmotivation korrelieren (von $r = .43$ bis $r = .93$). Zudem zeigten die einzelnen Skalen bis auf eine Ausnahme (Einstellungen) akzeptable bis sehr gute Werte für die interne Konsistenz und für die Split-Half-Koeffizienten nach der Spearman-Brown-Formel. Dass die Modellrechnung mithilfe von AMOS keinen adäquaten Modell-Fit für das vorgeschlagene Modell liefert, könnte, wie bereits diskutiert, daran liegen, dass der Zusammenhang zwischen den Skalen nicht linear ist, die Skalen an sich aber geeignet sind, die Testmotivation zu erfassen. So könnte es beispielsweise sein, dass die Skalen zwar die intendierten Konstrukte messen, diesen jedoch kein gemeinsames Konstrukt Testmotivation zugrunde liegt.

Die Ergebnisse der Modellrechnungen könnten sich jedoch auch durch eine andere Anordnung der Skalen untereinander verbessern. So könnten beispielsweise die beiden Skalen Erfolgserwartung und Aufgabenwert an oberster Stelle des Modells stehen. Diese beeinflussen dann, gemäß den Annahmen von Eccles und Kollegen (1983), das Leistungsverhalten im Test direkt, welches somit die Testmotivation widerspiegelt. Zudem gehen Eccles et al. (1983) davon aus, dass die Skalen Erfolgserwartung und Aufgabenwert ihrerseits durch die Einstellungen und das allgemeine Leistungsverhalten direkt beeinflusst werden, weshalb diese im Modell neben den eben genannten stehen und auf

diese gerichtet sein müssten. Ob das so vorgeschlagene Modell bessere und adäquate Modell-Fits liefert, kann im Rahmen dieser Arbeit nicht beantwortet werden.

5.2.3 *Intelligenz*

Auch in dieser Studie zeigten sich bei allen Subtests die höchsten Zusammenhänge genau bei den von Kreuzpointner (2010) zugewiesenen Komponenten. Darüber hinaus sprechen die Indizes der Strukturmodellierung für die Modellgültigkeit des LPS-neu (Version A: $\chi^2 = 46.79$ ($df = 40$; $p = .22$); RMSEA = .031; CFI = .979; Version B: $\chi^2 = 60.82$ ($df = 40$; $p = .02$); RMSEA = .053; CFI = .933). So kann zum einen gefolgert werden, dass es sich bei der allgemeinen Intelligenz um einen Faktor handelt, der sich aus mehreren Faktoren zusammensetzt und somit an oberster Ebene steht, zum anderen, dass sich die von Kreuzpointner (2010) vorgeschlagene Struktur des LPS-neu auch in dieser Studie replizieren ließ. Darüber hinaus unterscheiden sich die beiden Versionen A und B, den Gesamtscore betreffend, nicht signifikant voneinander ($t = .44$; $df = 359$; $p = .66$), in einzelnen Subtests und in der Stratum II kognitive Schnelligkeit hingegen schon.

Hinsichtlich des Geschlechts konnten, den Gesamtscore des LPS-neu betreffend, ebenfalls keine signifikanten Unterschiede gefunden werden. Dabei entspricht dieses Ergebnis den Forderungen an Intelligenztests, dass sich bei der Erfassung der allgemeinen Intelligenz keine Geschlechtsunterschiede einstellen sollten (Merz 1979; Neisser et al., 1996; Lubinski, Benbow & Morelock, 2000; zitiert nach Stapf 2003). Werden hingegen die einzelnen Subtests der Versionen A und B des LPS-neu betrachtet, so zeigen sich bei fünf signifikante Differenzen bezüglich des Geschlechts. Die Schülerinnen schneiden bei den Subtests 9 (8. Zeichen) und 10 (Zeilenvergleich) der Version A des LPS-neu etwas besser ab als die Schüler. Es scheint, dass die kognitive Schnelligkeit in dieser Stichprobe bei den Schülerinnen etwas höher ausgeprägt ist. Die Schüler erzielen hingegen sowohl im Subtest 4 (Zahlenfolgen) der Version A als auch im Subtest 1 (Allgemeinwissen) und 7 (Flächenzahl) der Version B bessere Ergebnisse und weisen somit eine höhere Ausprägung in der kristallinen und fluiden Intelligenz sowie in der visuellen Wahrnehmung auf. Entgegen dem zusammenfassenden Bericht von Neisser und Kollegen (1996) zeigen die Schülerinnen in dieser Studie keine besseren Werte bei den verbalen Aufgaben. Jedoch konnte sich das Befundmuster der Untersuchung von Colom und Lynn (2004) bestätigen. Auch in ihrer Studie erreichten die Schülerinnen bessere

Leistungen in der Verarbeitungsgeschwindigkeit, die Schüler hingegen erzielten unter anderem in numerischen und räumlichen Aufgaben bessere Ergebnisse.

Auch im Hinblick auf die untersuchten Schularten konnten signifikante Unterschiede bezüglich des Gesamtscores des LPS-neu nachgewiesen werden. So verhalten sich die gefundenen Mittelwerte entsprechend der Hierarchieeinteilung der einzelnen Schularten, wobei die höchsten Werte die Gymnasiasten/-innen und die niedrigsten die Hauptbeziehungsweise Mittelschüler/-innen erzielten. Es scheint also auch in dieser Studie die Höhe des angestrebten beziehungsweise erreichten Schulabschlusses mit der Intelligenz hoch positiv zu korrelieren (Amelang & Bartussek, 1997), da hier ebenfalls die Intelligenz mit dem angestrebten Schulabschluss steigt.

Keine signifikanten Differenzen ergaben sich jedoch bei der Berechnung des Zusammenhangs zwischen Alter und Gesamtscore des LPS-neu. Dabei lieferten zwar die 16-jährigen und 17-jährigen Schüler/-innen die höchsten Werte im Gesamtwert, jedoch unterscheiden sich diese nicht signifikant von den Ergebnissen der anderen Altersklassen. Auffällig ist, wenn auch nicht signifikant, dass der 19-jährige Schüler und die 18-jährigen Schülerinnen, also die ältesten Versuchspersonen in der teilnehmenden Stichprobe, die geringsten Werte im LPS-neu aufwiesen. Eine mögliche Ursache für diesen Befund könnte darin liegen, dass ebengenannte Testteilnehmer/-innen, verglichen mit dem Durchschnittsalter dieser Studie, bis zu vier Jahre älter sind und sie unter Umständen bereits mindestens eine Jahrgangsstufe wiederholen mussten. Das wiederum lässt entweder auf eine geringere allgemeine Intelligenz schließen oder könnte auf einen Mangel an Leistungsbereitschaft, wie diese auch im LPS-neu aufzubringen ist, hinweisen. Zumindest in dieser Studie wies die Gruppe der 18-jährigen Mädchen den niedrigsten Mittelwert der gesamten Testmotivation ($M = 154.21$; $SD = 19.67$) auf. So könnte die mangelnde Leistungsbereitschaft zu den geringen Werten im LPS-neu geführt haben.

5.2.4 Zusammenhänge

5.2.4.1 Leistungsmotiv und Testmotivation

Ausgehend davon, dass die Testmotivation durch die Leistungsmotive beeinflusst wird, konnte gezeigt werden, dass diese am höchsten bei den Versuchspersonen ist, bei denen die Hoffnung auf Erfolg überwiegt. Die Hoffnung auf Erfolg beeinflusst also die Motivation für einen bestimmten Test positiv. Es zeigte sich außerdem, dass Personen, bei denen sowohl die Hoffnung auf Erfolg als auch die Furcht vor Misserfolg hoch ausge-

prägt sind, am höchsten testmotiviert waren. So haben diese beiden Komponenten bei hoher Ausprägung einen starken Einfluss auf die Testmotivation und somit wiederum auf die Leistungsbereitschaft, in einem Test das Beste zu geben. Entgegen der Annahme lag bei Versuchspersonen mit überwiegend LL ebenfalls eine ausreichende Testmotivation vor. Es könnte demnach sein, dass Personen mit einer starken FE in ihrer Angst vor Misserfolg beziehungsweise Personen mit einer hohen HM in ihrer Hoffnung auf Erfolg keine mittlere, sondern eine hohe Ausprägung aufweisen. Somit wären sie mit den Versuchspersonen, bei denen das Leistungsmotiv HH überwiegt, vergleichbar. Deshalb könnte auch hier der Schluss gezogen werden, dass eine hohe Ausprägung in der Hoffnung auf Erfolg und in der Furcht vor Misserfolg die Testmotivation positiv beeinflusst. Die geringste Testmotivation zeigte sich bei den Versuchspersonen mit hoher Angst vor Misserfolg bei gleichzeitiger geringer Hoffnung auf Erfolg. Dementsprechend könnte das schlechtere Abschneiden von Misserfolgsmotivierten in den Tests nicht nur daran liegen, dass die zur Lösung der Aufgaben notwendigen Kognitionen mit der Furcht in Tests interferieren (Sarason, 1963), sondern auch daran, dass die Furcht vor Misserfolg bereits die Testmotivation negativ beeinflusst und dieses wiederum zu schlechteren Ergebnissen führt.

Betrachtet man nun noch die rotierte Komponentenmatrix der Leistungsmotivcluster (siehe Tabelle 4.3; Kapitel 4.1), so fällt auf, dass das Leistungsmotiv LL, zumindest beim zweiten Faktor – dem unabhängigen Konstrukt FM –, zwar eine geringere, aber dennoch vergleichbare Ausprägung aufweist wie das Motiv LH. Diese beiden Leistungsmotive sind somit durch eine hohe FM gekennzeichnet. Zusätzlich ergab nach signifikanten Mittelwertsunterschieden der Scheffé Post-Hoc-Test für die Testmotivation der vier Leistungsmotivtypen (siehe Tabelle 4.7; Kapitel 4.4.1), dass die Testmotivation des Leistungsmotivclusters LL nicht niedriger ist als die der Cluster HL und HH, aber auch nicht höher als die des Clusters LH. Die aber im Vergleich zu HL und HH geringere Testmotivation könnte hier durch eine hohe FM bei gleichzeitig geringer HE begründet sein. Da die FM jedoch geringer ausgeprägt ist als bei LH, konnten die Versuchspersonen dieses Leistungsmotivcluster dennoch eine ausreichende Testmotivation aufbringen.

Es bleibt festzuhalten, dass die Leistungsmotivcluster HH, HL und LL eine ausreichende Testmotivation für die Bearbeitung eines Leistungstests aufweisen, die Versuchspersonen mit überwiegend LH hingegen nicht. Die vier Leistungsmotivtypen unterscheiden sich zwar signifikant in ihrer Höhe der Testmotivation voneinander, jedoch

lässt sich bei Kenntnis des Leistungsmotivtyps nicht eindeutig auf die vorhandene Testmotivation schließen.

Untersucht wurde im Rahmen dieser Studie nicht, ob sich die verschiedenen Leistungsmotivtypen in ihren Intelligenzleistungen statistisch signifikant unterscheiden, da angenommen wurde, dass dieser Zusammenhang über die Testmotivation vermittelt wird.

5.2.4.2 Testmotivation und Leistungen in einem Leistungstest

Die Untersuchung zeigte, dass die Testmotivation einen zwar geringen, aber positiv signifikanten Einfluss auf die Leistungen im Leistungsprüfsystem-neu ausübt ($r = .16$; $p = .00$). Zudem wurde der Einfluss der Testmotivation, die vor dem Test erhoben wurde, getrennt von der Testmotivation, die nach dem Test erfasst wurde, auf die Leistungen im LPS-neu berechnet. Dabei fielen die getrennt berechneten Strukturmodellrechnungen zusammen mit dem LPS-neu etwas besser aus, was dafür spricht, die Testmotivation vor dem Test getrennt von der Testmotivation nach dem Test zu betrachten. Darüber hinaus scheint es so zu sein, dass die Schüler/-innen ihre Testmotivation nach dem Test besser einschätzen konnten als vor dem Test. Zumindest sind hier die Zusammenhänge mit den Leistungen des LPS-neu Gesamtwertes deutlich höher gewesen (vorher: $r = .07$; $p = .18$; nachher: $r = .23$; $p = .00$). Folglich hat zumindest in dieser Studie die wahrgenommene Leistung im LPS-neu auf die Testmotivation (Sanchez et al., 2000) einen höheren Einfluss. Die Testmotivation, die vor der Bearbeitung des Leistungstests vorhanden war und somit ohne Wissen der dann tatsächlich zu erbringenden Leistungen erhoben wurde, hatte hingegen weniger Einfluss.

Aufgrund des zwar geringen, aber signifikant positiven Zusammenhangs zwischen Testmotivation und Intelligenz konnte bestätigt werden, dass der Leistungsanstieg im LPS-neu durch eine höhere Testmotivation begründet ist. Zudem sollte wegen kognitiver Grenzen die Korrelation bei hoher Testmotivation gegen Null tendieren ($r = .15$; $p = .18$), was ebenfalls bestätigt werden konnte. Jedoch führt in der Tendenz eine höhere Testmotivation auch zu höheren Leistungen im LPS-neu Gesamtwert. Das spricht gegen die vorher aufgestellte Hypothese. So kann es sein, dass sich die Versuchspersonen mit hoher Testmotivation noch stärker konzentriert haben, aufmerksamer waren und diese Aufmerksamkeit auch über einen längeren Zeitraum aufrechterhalten haben, was gerade

beim Leistungsprüfsystem-neu von besonderer Bedeutung ist und somit zu einer Leistungssteigerung – ohne Anfälligkeit für einen Deckeneffekt – geführt hat.

Ebenfalls konnte in dieser Studie das Ergebnis von Rindermann (2006; zitiert nach Jacobs, 2007) bestätigt werden, nach welchem für ein Testergebnis in Leistungstests die Intelligenz verantwortlich ist, insofern eine ausreichende Motivation für die Bearbeitung dieses Tests vorliegt. So wurde in dieser Arbeit ein Zusammenhang von $r = .08$ ($p = .19$) zwischen der ausreichenden Testmotivation und der allgemeinen Intelligenz des LPS-neu berechnet. Aufgrund der nicht vorhandenen Signifikanz dieser Korrelation kann geschlussfolgert werden, dass bei ausreichender Testmotivation die Intelligenz für das Testergebnis im LPS-neu verantwortlich ist. Beim Vorhandensein von hoher Testmotivation zeigte sich hingegen tendenziell ein positiver und größerer Einfluss auf die Leistungen im LPS-neu. Um dementsprechend bessere Leistungen zu erzielen, wäre es empfehlenswert, die Schüler/-innen durch geeignete Methoden nicht nur ausreichend für einen Test zu motivieren. Es könnte jedoch auch sein, dass gerade intelligentere Schüler/-innen in der Tendenz auch eine höhere Testmotivation aufweisen, da sie bereits ein positives Leistungsergebnis erwarten, weshalb sie wiederum motivierter bei der Bearbeitung des Tests sind. Gemeinsam haben diese beiden Ursachen dann zur Folge, dass tatsächlich höhere Leistungen erzielt werden.

5.2.4.3 Schulleistung, besuchte Schulart bzw. Jahrgangsstufe und Intelligenz

In dieser Studie zeigten sich nur geringe Zusammenhänge zwischen guten Schüler/-innen und einer höheren allgemeinen Intelligenz. Die höchsten Werte zwischen der Schulleistung und dem LPS-neu Gesamtscore wurden hier an der Grund- und Mittelschule gefunden ($r = -.44$; $p = .01$). Die Berechnungen der anderen Schularten lieferten für diesen Zusammenhang keine Signifikanz. Demzufolge scheint die allgemeine Intelligenz, welche nach Rost (2009) als bester Einzelprädiktor für die Schulleistung gesehen werden kann, zumindest in dieser Studie kein besonders geeigneter Prädiktor zu sein. Darüber hinaus wurden auch die Haupt- und Nebenfachnoten bezüglich ihres Zusammenhangs mit der allgemeinen Intelligenz untersucht. Es konnte hier ebenfalls nur an der Grund- und Mittelschule eine signifikante Korrelation zwischen den Hauptfachnoten – Mathematik, Deutsch, Englisch – und der allgemeinen Intelligenz gefunden werden ($r = .59$; $p = .00$). Die Ergebnisse der anderen Schularten sowie die der Zusammenhänge mit den Nebenfachnoten waren durchweg nicht signifikant. Ferner wurde

mithilfe multipler Regressionsanalysen überprüft, ob sich entsprechend einiger bisheriger Untersuchungen beziehungsweise Metaanalysen (Funke & Vaterrodt-Plünnecke, 2004; Kühn, 1987) die Zensuren in den Hauptfächern besser als Prädiktoren für die allgemeine Intelligenz eignen als die Nebenfachnoten. Auch das konnte nur teilweise bestätigt werden. Lediglich an der Grund- und Mittelschule und an der Mädchenrealschule sind die Hauptfächer als Prädiktoren besser geeignet. Die Ergebnisse von Wechslers (1964) Untersuchung konnten hingegen in dieser Studie repliziert werden. So hat die besuchte Schulart einen Einfluss auf die Leistungen im LPS-neu, wobei die Grund- und Mittelschüler/-innen die niedrigsten und die Schüler/-innen des Gymnasiums die höchsten Werte im LPS-neu Gesamtscore erreichten.

Der eher geringe Zusammenhang zwischen den Leistungen im LPS-neu und den Schulleistungen beziehungsweise den Haupt-/Nebenfachnoten könnte daran liegen, dass dieser Test keine schulnahen Leistungen erfasst, sondern vielmehr die allgemeine Leistungsfähigkeit, die Leistungsbereitschaft und die Leistungsmotivation anspricht. Es sind zwar kognitive Lernvoraussetzungen, wie beispielsweise die Intelligenz und fachliche Fähigkeiten, wichtige Prädiktoren für die Schulleistung, aber diese weisen meist nur dann höhere Zusammenhänge bei Intelligenztests auf, wenn die Tests vor allem verbal stark ausgeprägt sind (Helmke, 1992). Zu diesen Tests zählt das LPS-neu mit lediglich zwei Subtests, die auch die verbalen Fähigkeiten erfassen, nicht. So scheint es, dass gerade in verbal ausgerichteten Tests das Wissen im Vergleich zur Intelligenz unterschätzt wird. Darüber hinaus konnte in vielen Untersuchungen ein größerer Einfluss auf die Leistungen in Intelligenztests durch Wissen als durch Intelligenz gefunden werden (Helmke, 1992). Auch bei Simons et al. korrelierten die Leistungsunterschiede stärker mit dem Vorwissen als mit der Intelligenz (Simons et al., 1975; zitiert nach Lukesch, 2006). Doch sollte dieser Vorkenntnisstand nicht unabhängig von der Intelligenz betrachtet werden. Intelligenz leistet einen entscheidenden Beitrag dafür, wie viel Wissen sich eine Person aneignen und anwenden kann beziehungsweise wie schnell das Wissen wieder vergessen wird (Gage, 1986). Somit ließe sich auch in dieser Studie der niedrige Zusammenhang zwischen Schulleistung und Intelligenz durch einen großen Einfluss der Vorkenntnisse auf die Lösung der Aufgaben erklären. Dieser Schluss erhält zudem dadurch Unterstützung, dass die Ergebnisse des LPS-neu Gesamtwertes der Grund- und Mittelschule deutlich höhere Zusammenhänge mit der Schulleistung aufweisen als die der Realschulen beziehungsweise des Gymnasiums. Werden die Realschule und das Gymnasium als höheres Ausbildungsniveau betrachtet, in welchem sich mehr Wissen

aneignen lässt, so ist hier der Einfluss auf die Schulleistung eher durch den Vorkenntnisstand als durch die Intelligenz geprägt (Helmke & Weinert, 1997). Unterstützung findet diese Annahme ebenfalls durch die Studie von Köller und Karim (2001), in welcher sich die Intelligenzentwicklung am Gymnasium, bei gleicher Ausgangsintelligenz, als am günstigsten erwies (zitiert nach Köller & Baumert, 2002). Schlussfolgernd werden die Lernfähigkeit und die Intelligenz vom schulischen Milieu geprägt, welche wiederum Ergebnis und Voraussetzung schulischer Lernprozesse sind. Das heißt, auf einem höheren Schulniveau ist der Zusammenhang zwischen Schulleistung und Vorkenntnisstand aufgrund von Lernfähigkeit, Leistungsanstrengung und bisheriger Wissensaneignung höher. Die Eltern haben daran sicher ebenfalls einen Anteil, denn bei eher schlechteren Schüler/-innen einer höheren Schule erwarten sie eine größere Leistungsanstrengung und somit bessere Leistungen. Wobei die Eltern ihre Kinder durch Nachhilfe oder Hausaufgabenkontrolle unterstützen und sie zudem an die Fähigkeiten ihrer Kinder glauben, was sich wiederum positiv auf die Schulleistung, aber weniger auf die allgemeine Intelligenz auswirkt. Dabei erwies sich vor allem die Zuversicht der Eltern „in die Leistungsfähigkeit und Entwicklungsfähigkeit des Kindes sowie prozessorientierte fachliche Unterstützung statt produktorientiertem Engagement“ (Helmke, Schrader & Lehneis-Klepper, 1991; zitiert nach Helmke, 1992) als leistungsförderlich. So kann auch hier der Schluss gezogen werden, dass durch eine erhöhte Leistungsanstrengung zwar das Wissen erhöht wird und somit der Einfluss auf die Schulleistung steigt, dieses jedoch nicht zwingend zu einer höheren Intelligenz beiträgt, weshalb dieser Einfluss geringer ausfällt.

Der Unterschied in den untersuchten Schularten kann aber auch, wie bereits einige Studien bestätigen konnten, aufgrund räumlicher Bedingungen zustande gekommen sein. So wiesen beispielsweise bei Coleman (1966) und bei Wheeler (1942) Stadtkinder durch eine bessere Infrastruktur und vorteilhaftere Schulausstattung höhere Intelligenzwerte als Landkinder auf (zitiert nach Gage, 1986). Da die Mittelschüler/-innen in dieser Studie als Einzige eine Schule auf dem Land besuchten, könnte die räumliche Bedingung ebenfalls ein weiterer Aspekt der niedrigeren Ergebnisse im LPS-neu sein.

Zudem erbrachten die Schüler/-innen der 10. Jahrgangsstufe im Mittel signifikant bessere Ergebnisse als diejenigen, die eine Klassenstufe unter ihnen waren. Diese Unterschiede konnten sowohl auf das Alter als auch auf die Jahrgangsstufe zurückgeführt werden. So weisen unterschiedliche Altersklassen unterschiedliche Intelligenztestergebnisse auf und auch die Jahrgangsstufen zeigen voneinander signifikant verschiedene

Ergebnisse. Im Durchschnitt waren die Schüler/-innen der 9. Jahrgangsstufe ein Jahr jünger (9. Klasse $M = 15.02$ ($SD = .78$); 10. Klasse: $M = 16.07$ ($SD = 1.00$)). Die Ergebnisse deuten somit darauf hin, dass die Intelligenzentwicklung bei den jüngeren Versuchspersonen noch nicht abgeschlossen ist. Nach Oerter (1975) steigt die Intelligenz in den ersten 15 Jahren stark an, stabilisiert sich schnell, erreicht aber erst ungefähr im 20. Lebensjahr den Höhepunkt. Ein weiterer Grund für die höheren Leistungen im LPS-neu der 10. Jahrgangsstufe könnte auch die Schulbesuchsdauer sein. Schüler/-innen höherer Klassen konnten sich bereits mehr Wissen aneignen und sind mit Leistungssituationen, wie diese auch im LPS-neu vorherrschen, bereits vertrauter. Auch Ceci (1991) konnte eine sehr deutliche Korrelation zwischen der Schuldauer und den Intelligenzmaßen finden (zitiert nach Lukesch, 2006).

Ebenfalls wurde geprüft, ob in dieser Studie ein Rückgang der Korrelation zwischen Schulleistung und Intelligenz mit steigendem Alter und dem Ausbildungsniveau vorliegt. Den Berechnungen zufolge konnte keine signifikante Abnahme des Zusammenhangs zwischen Schulleistung und Intelligenz bei gleichzeitig steigendem Alter gefunden werden. Betrachtet man hingegen die Jahrgangsstufe 10 als höheres Ausbildungsniveau im Vergleich zur 9. Jahrgangsstufe, so konnte hier ein Absinken des Zusammenhangs zwischen Schulleistung und Intelligenztestleistungen festgestellt werden. Jedoch lieferte nur das Gymnasium signifikante Ergebnisse. Demzufolge scheint auch hier vor allem bei Schüler/-innen, die das Gymnasium besuchen, das Vorwissen eine bedeutende Rolle zu spielen (Helmke & Weinert, 1997). So ist gerade bei diesem Schultyp der Schulerfolg in den höheren Klassen beziehungsweise am Ende einer Schullaufbahn vorrangig vom Wissen abhängig, welches über die Schuljahre hinweg aufgebaut wurde. An den anderen Schularten scheint der Zusammenhang zwischen Jahrgangsstufe und Schulleistung mit der allgemeinen Intelligenz hingegen weniger Gewicht zu haben.

5.3 Nachträgliche Hypothesen

Die nachfolgenden Hypothesen resultierten aus den vorgefundenen Ergebnissen und der anschließenden Diskussion. Zunächst sollten diejenigen Schüler/-innen, die eine Klasse im vorherigen Schuljahr wiederholen mussten, hinsichtlich ihrer Testmotivation genauer untersucht werden.

Es wurde errechnet, dass die Wiederholer schlechtere Ergebnisse im LPS-neu ablieferten als diejenigen Schüler/-innen, die keine Klasse wiederholen mussten. Nun stellt

sich die Frage, ob dieser Umstand damit zusammenhängt, dass bei ihnen auch eine signifikant niedrigere Testmotivation vorliegt, da sie geringe Leistungserfolge vorweisen und somit geringere Anstrengung aufbringen. Aus diesem Grund sollte die untenstehende Hypothese 14 überprüft werden.

H14: Diejenigen Schüler/-innen, die eine Klasse wiederholen müssen, weisen in Leistungstests wie dem LPS-neu signifikant niedrigere Testmotivation auf.

Werden nun die Mittelwerte der gesamten Testmotivation von den verbleibenden 28 Wiederholern ($M = 161.40$; $SD = 28.72$) mit den 267 Schüler/-innen verglichen, die keine Klasse wiederholen mussten ($M = 167.22$; $SD = 25.11$), so weisen diese geringere Werte auf. Doch sind diese signifikant voneinander verschieden? Nach Berechnung eines Welch-Tests ($t(31.48) = 1.06$; $p = .31$) wird klar, dass diese beiden Gruppen sich statistisch nicht signifikant in der Höhe ihrer Testmotivation unterscheiden. Die geringeren Ergebnisse im LPS-neu liegen daher nicht an einer niedrigeren Testmotivation.

Eine weitere mögliche Hypothese ergab sich aus dem Verhalten der Mittelschüler/-innen während der Testung. So wäre es nachträglich interessant zu prüfen, ob diese sich nicht nur signifikant in ihren Intelligenzleistungen, sondern auch in der Höhe ihrer Testmotivation von den anderen Schularten unterscheiden. Die schlechtere Leistung der Mittelschüler/-innen im LPS-neu könnte sich aus einer mangelnden Testmotivation ergeben. Aus diesem Grund wird folgende Hypothese aufgestellt:

H15: Mittel- beziehungsweise Hauptschüler/-innen weisen eine geringere Testmotivation auf als Schüler/-innen anderer Schularten.

Werden die Schulen einzeln nach der Höhe ihrer gesamten Testmotivation betrachtet (siehe Tabelle 5.1), so wird deutlich, dass sie sich kaum in ihren Mittelwerten unterscheiden. Auch die Minimum- und Maximumwerte sind fast identisch.

Tabelle 5.1

Mittelwerte, Minimum- und Maximumwerte der gesamten Testmotivation, getrennt nach Schulen

	<i>M (SD)</i>	<i>Min</i>	<i>Max</i>
Mittelschule	165.31 (28.28)	113.00	220.00
Mädchenrealschule	165.19 (24.23)	116.00	217.00
Knabenrealschule	168.33 (24.34)	102.47	221.00
Gymnasium	166.39 (26.75)	81.00	232.00

Anmerkungen. *M* = Mittelwert; *SD* = Standardabweichung

Auch eine univariate Varianzanalyse mit der gesamten Testmotivation als abhängige Variable zwischen den verschiedenen Schulen brachte kein signifikantes Ergebnis ($F(3) = .24$; $p = .87$; $\eta^2 = .00$). Somit kann die Hypothese 15 an dieser Stelle nicht bestätigt werden. Grund- und Mittelschüler/-innen zeigen nicht aufgrund geringerer Testmotivation schlechtere Leistungen im LPS-neu. Vielmehr liegt es daran, dass die Schüler/-innen in den verschiedenen Schularten unterschiedlich intelligent sind.

5.4 Ausblick

Mithilfe dieser Untersuchung konnte die Testmotivation, also die Motivation, in einer bestimmten Situation oder in einem Test erfolgreich zu sein (Eklöf, 2008), genauer untersucht werden. Es zeigte sich, dass die Testmotivation zwar nur geringe, aber signifikant positive Auswirkungen auf die Leistung in einem Leistungstest hat, der ohne Konsequenzen für die Versuchspersonen bleibt.

Deshalb sollte die Testmotivation nicht nur in Gruppen-, sondern auch in Einzeluntersuchungen gemessen werden, da die sonst erhaltenen Daten zu ungültigen Testergebnissen und somit zu falschen Schlussfolgerungen führen können (Merten & Brockhaus, 2004). Wird diese Variable in der Forschung nicht berücksichtigt, können irreführende Ergebnissen resultieren, was wiederum die psychologische Theorienbildung beeinflussen kann.

Für weiterführende Untersuchungen wäre es unter anderem interessant zu überprüfen, ob sich ein höherer Zusammenhang erzielen lässt, wenn einzelne Items des Fragebogens geändert werden beziehungsweise die Struktur des Modells verändert wird, wie dies bereits in der Diskussion vorgeschlagen wurde. Zudem könnte dabei auch der größere Einfluss der Testmotivation, die nach dem Test erhoben wurde, genauer untersucht

werden. Weitere Untersuchungen könnten sich ferner darauf konzentrieren, inwieweit sich Versuchspersonen, bei denen es um Testungen ohne Konsequenzen geht, in ihrer Motivation beispielsweise von Patienten unterscheiden, die durch die Erkenntnisse solcher Tests vielleicht sogar gesund werden können. Dieser Punkt wäre gerade für Schulleistungsstudien – die bekannteste unter ihnen ist die PISA-Studie – von besonderer Bedeutung, da diese Tests vor allem darin kritisiert werden, dass sie eher die Motivation als die Leistungsfähigkeit der Schüler/-innen widerspiegeln. Dabei sollte die Testmotivation gerade wegen ihres Einflusses alle Bereiche der Psychologie betreffen, wie etwa die angewandte Psychologie oder die grundlagenorientierte Forschung (Green, 2004) und hier bereits bei der Gestaltung von Tests und bei den Testinstruktionen berücksichtigt werden. Zudem kann dadurch die Möglichkeit geschaffen werden, mehr über den Mangel an Testmotivation zu erfahren und gegebenenfalls diesen Mangel durch geeignete Verfahren zu minimieren. Darüber hinaus könnte diese Forschung dazu beitragen, genaueres Wissen über die Testmotivation zu erlangen, um beispielsweise zu erfahren, welche Bedingungen für eine ausreichende Testmotivation gegeben sein müssen. Nur so kann sichergestellt werden, dass ein Test nicht nur den Mangel an Testmotivation erfasst, sondern das tatsächlich zu messende Konstrukt bestimmt.

Auch wurden die Auswirkungen der Testmotivation auf die Leistungen in einem Leistungstest in dieser Studie nur an Jugendlichen im Alter von 14 bis 19 Jahren getestet. Es wäre also interessant zu überprüfen, ob sich die erzielten Untersuchungsergebnisse auch bei einer Erwachsenenstichprobe replizieren lassen. Hier könnte es beispielsweise sogar sein, dass die Skala Einstellungen des Testmotivationsfragebogens bessere Ergebnisse liefert und somit die gesamte Testmotivation höhere Zusammenhänge aufweist. Dazu müssten jedoch die Items des Fragebogens auf diese Altersgruppe angepasst werden.

6. Zusammenfassung

In der Psychologie werden die meisten Erkenntnisse durch Testungen bzw. Untersuchungen mit Patienten oder anderen Versuchspersonen ermöglicht. Sie alle nehmen freiwillig teil. Doch wie sieht es mit der Motivation der Testpersonen aus? Schulleistungsstudien sorgen in Deutschland immer wieder für große Aufregung, nicht nur die bekannteste unter ihnen, die PISA-Studie. Kritisiert wird bei diesen Tests häufig, dass sie eher die Motivation als die Leistungsfähigkeit der Schüler/-innen widerspiegeln.

In der vorliegenden Studie war es das Ziel, den Einfluss der Testmotivation auf die Leistungen in einem Leistungstest zu explorieren. Dabei wurden zunächst die grundlegenden Leistungsmotive mithilfe des Regensburger Leistungs-Motiv-Inventars für Kinder und Jugendliche (RLMI-K/J; Lukesch, in Vorb.) und des Leistungsmotivationsinventars für Kinder und Jugendliche (kurz; AMG-S K-J; Schmalt, 1999; 2005) erfasst und in ein gemeinsames Modell – das quadripolare Modell der Leistungsmotivation (Covington & Omelich, 1991; Covington & Roberts, 1994) – integriert. Anschließend wurden diese mit der situationsspezifischen Testmotivation, die mittels eines selbst entwickelten Fragebogens erhoben wurde, basierend auf der Erwartungs-Wert-Theorie der Leistungsmotivation von Eccles und Kollegen (1983), in Verbindung gebracht. Darüber hinaus wurden die besuchte Schulart und die Schulleistungen erhoben sowie deren Einfluss als Prädiktor für die allgemeine Intelligenz untersucht. Die überarbeitete Version des Leistungsprüfsystems (LPS-neu) von Kreuzpointner (2010) diente dabei als Leistungstest. Die Stichprobe dieser Untersuchung setzte sich aus insgesamt 361 ($w = 154$; $m = 207$) Schülern und Schülerinnen zusammen, wobei sich jene auf vier staatliche bayerische Schulen (ein Gymnasium, zwei Realschulen sowie eine Grund- und Mittelschule) verteilten. Die Probanden waren zum Zeitpunkt der Untersuchung zwischen 14 und 19 Jahre alt.

Es ergab sich ein signifikant positiver Einfluss der Testmotivation auf die Leistungen im LPS-neu. Dementsprechend konnte ein Leistungsanstieg im LPS-neu tendenziell durch eine höhere Testmotivation begründet werden. Im hohen Testmotivationsbereich wurde die Korrelation hingegen nicht signifikant, weshalb hier, wie erwartet, ausschließlich die Intelligenz für das Testergebnis im LPS-neu verantwortlich war.

Das quadripolare Modell erwies sich als geeignet, um die Leistungsmotive der verwendeten Fragebögen zu integrieren. Die dabei entstandenen vier Leistungsmotivcluster unterschieden sich signifikant in ihrer Testmotivation, wobei bei Kenntnis des Leis-

tungsmotivs nicht unmittelbar auf die Höhe der vorliegenden Testmotivation geschlossen werden konnte.

Die von Kreuzpointner (2010) formulierte Struktur des LPS-neu ließ sich replizieren. Darüber hinaus konnte jedoch nur ein geringer Zusammenhang zwischen dem LPS-neu Gesamtscore und der Schulleistung gefunden werden. Auch erreichten die Hauptfachnoten, im Vergleich zu den Nebenfachnoten, nicht an jeder untersuchten Schule eine bessere Vorhersagekraft für die allgemeine Intelligenz. Die besuchte Schulart erwies sich jedoch als einflussreich auf die Leistungen im LPS-neu, wobei die Ergebnisse entsprechend der Hierarchie der Bildungsstufen ausfielen.

7. Literaturverzeichnis

- Amelang, M. (1976). Validierung von Anforderungsprofilen für das Studium der Medizin, Zahnmedizin, Pharmazie und Psychologie. *Psychologie in Erziehung und Unterricht*, 23, 259-273.
- Amelang, M. & Bartussek, D. (1997). *Differentielle Psychologie und Persönlichkeitsforschung* (4. überarbeitete und erweiterte Auflage). Stuttgart: Kohlhammer.
- Amelang, M. & Zielinski, W. (2002). *Psychologische Diagnostik und Intervention* (3. korrigierte, aktualisierte und überarbeitete Auflage). Berlin: Springer.
- Amthauer, R., Brocke, B., Liepmann, D. & Beauducel, A. (1999, 2001). *Intelligenz-Struktur-Test 2000 (I-S-T 2000)*. Göttingen: Hogrefe.
- Atkinson, J. W. (1957). Motivational determinants of risk-taking behaviour. *Psychological Review*, 64, 359-372.
- Atkinson, J. W. (1958). Towards experimental analysis of human motivation in terms of motives, expectancies, and incentives. In J. W. Atkinson (Hrsg.), *Motives in fantasy, action and society. A method of assessment and study* (S. 288-305). Princeton: Van Nostrand.
- Atkinson, J. W. (1975). *Einführung in die Motivationsforschung*. Stuttgart: Klett. (Original erschienen 1964: An introduction to motivation).
- Atkinson, J. W. & Litwin, G. H. (1960). Achievement motive and test anxiety conceived as motive to approach success and motive to avoid failure. *Journal of Abnormal and Social Psychology*, 60, 52-63.
- Arvey, R. D., Strickland, W., Drauden, G. & Martin, C. (1990). MOTIVATIONAL COMPONENTS OF TEST TAKING. *Personnel Psychology*, 23, 695-716.
- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2008): *Multivariate Analysemethoden* (12. Auflage). Berlin, Heidelberg, New York: Springer.
- Baumert, J. & Demmrich, A. (2001). Test motivation in the assessment of student skills: The effects of incentives on motivation and performance. *European Journal of Psychology of Education*, 16, 441-462.
- Berk, L. E., Schönpflug, U., Petersen, K. & Aralikatti, E. (2011). *Entwicklungspsychologie* (5. aktualisierte Auflage). München: Pearson Studium.
- Bjorklund, D. F. & Schneider, W. (2006). Ursprung, Veränderung und Stabilität der Intelligenz im Kindesalter: Entwicklungspsychologische Perspektiven. In W. Schneider & B. Sodian (Hrsg.), *Enzyklopädie der Psychologie, Serie V Entwicklungspsychologie, Band 2: Kognitive Entwicklung* (S. 769-821). Göttingen: Hogrefe.
- Bortz, J. (1999). *Statistik für Sozialwissenschaftler* (5. Auflage). Berlin: Springer.

- Bossong, B. (1985). Testmotivation – Geben Sie wirklich Ihr Bestes? In F.- J. Hehl (Hrsg.), *12 Kongress für Angewandte Psychologie, 21. - 25. September 1983, Universität Düsseldorf. Band 3: Diagnostik und Evaluation bei betrieblichen, politischen und juristischen Entscheidungen* (S. 88-100). Bonn: Deutscher Psychologen-Verlag.
- Brambring, M. (1983). Spezielle Eignungsdiagnostik. In K. J. Groffmann & L. Michel (Hrsg.), *Enzyklopädie der Psychologie: Themenbereich B, Methodologie und Methoden: Serie 2 Psychologische Diagnostik, Band 2 Intelligenz- und Leistungsdiagnostik* (S. 414-481). Göttingen: Hogrefe.
- Brunstein, J. & Heckhausen, H. (2006). Leistungsmotivation. In J. Heckhausen & H. Heckhausen (Hrsg.), *Motivation und Handeln* (3. überarbeitete und aktualisierte Auflage), (S. 143-191). Heidelberg: Springer Medizin Verlag.
- Buch, S. (2007). Strukturgleichungsmodelle – Ein einführender Überblick. *ESCP-EAP Working Paper No. 29*. (Online verfügbar: http://www.escp-eap.eu/uploads/media/Buch_Strukturgleichungsmodelle_ESCP_EAP_WP29_04.pdf)
- Carroll, J. B. (1993). *Human cognitive abilities. A survey of factor-analytic studies*. Cambridge: Cambridge University Press.
- Carroll, J. B. (1996). A Three-Stratum Theory of Intelligence: Spearman's contribution. In I. Dennis & P. Tapsfield (Eds.), *Human abilities* (S. 1-19). New Jersey: Lawrence Erlbaum Associates.
- Cattell, J. M. (1890). Mental tests and measurements. *Mind*, 15, 373-380.
- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54, 1-22.
- Cattell, R. B. (1971). *Grundintelligenztest CFT 3 Skala 3*. Braunschweig: Westermann.
- Chan, D., Schmitt, N., Sacco, J. M. & DeShon, R. P. (1998). Understanding pretest and posttest reactions to cognitive ability and personality tests. *Journal of Applied Psychology*, 83, 471-485.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (Second Edition). Hillsdale, New York: Erlbaum Associates.
- Cole, J. S. (2007). *MOTIVATION TO DO WELL ON LOW-STAKES TESTS*. University of Missouri-Columbia. (Online verfügbar: <http://proquest.umi.com/pqdlink?did=1564005121&Fmt=7&clientId=79356&RQT=309&VName=PQD>)
- Colom, R. & Lynn, R. (2004). Testing the developmental theory of sex differences in intelligence on 12-18 year olds. *Personality and Individual Differences*, 36, 75-82.
- Conrad, W. (1983). Intelligenzdiagnostik. In K.-J. Groffmann & L. Michel (Hrsg.), *Enzyklopädie der Psychologie: Themenbereich B: Methodologie und Methoden. Serie 2: Psychologische Diagnostik, Band 2 Intelligenz- und Leistungsdiagnostik* (S. 104-201). Göttingen: Hogrefe

- Covington, M. V. & Omelich, C. L. (1991). Need achievement revisited: Verification of Atkinson's original 2 x 2 model. In C. D. Spielberger, I. G. Sarason, Z. Kulcsar, & G. L. Van Heck (Eds.), *Stress and emotion: Anxiety, anger and curiosity* (S. 85–105). Washington, DC: Hemisphere.
- Covington, M. V. & Roberts, B. W. (1994). Self-worth and college achievement: Motivational and personality correlates. In P. R. Pintrich, D. R. Brown & C. E. Weinstein (Eds.), *Student motivation, cognition, and learning* (S. 157–188). Hillsdale, NJ: Erlbaum.
- Daseking, M. Janke, N. & Petermann, F. (2006). Intelligenzdiagnostik. *Monatsschrift Kinderheilkunde*, 154, 314-319.
- Daumenlang, K. (1990). Intelligenztests. In W. Sarges (Hrsg.), *Management-Diagnostik* (3. Auflage) (S. 540-548). Göttingen: Hogrefe.
- Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L. & Migdley, C. (1983). Expectancies, values and academic behaviors. In J. T. Spence (Ed.), *Achievement and achievement motives: Psychological and sociological approaches* (S. 75-146). San Francisco: Freeman.
- Eccles, J.S., Wigfield, A. & Schiefele, U. (1998). Motivation to succeed. In W. Damon (Series Ed.) and N. Eisenberg (Vol. Ed.), *Handbook of child psychology* (5th Edition, Vol. III), (S. 1017-1095). New York: Wiley.
- Eccles, J. S. & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53, 109–132.
- Edelmann, W. (2000). *Lernpsychologie* (6., vollständig überarbeitete Auflage). Weinheim: Beltz.
- Eiseler, N. (2010). *Der Einfluss von Misserfolgsstress auf die Intelligenztestleistung in Abhängigkeit von dem Leistungsmotiv*. Universität Regensburg: Unveröffentlichte Diplomarbeit.
- Eklöf, H. (2008). Test-taking motivation on low-stakes tests: A Swedish TIMSS 2003 example. In *Issues and Methodologies in Large-Scale Assessments*, IERI Monograph Series, Vol. 1. Hamburg: IEA-ETS Research Institute.
- Eklöf, H. (2006). Development and Validation of Scores From an Instrument Measuring Student Test-Taking Motivation. *Educational and Psychological Measurement*, 66, 643-656.
- French, E. G. (1958). Effects of the interaction of motivation and feedback on task performance. In J. W. Atkinson (Ed.), *Motives in fantasy, action and society. A method of assessment and study* (S. 400-408). Princeton: Van Nostrand.
- French, E. G. & Thomas, H. T. (1958). The relation of achievement motivation to problem solving effectiveness. *Journal of Abnormal and Social Psychology*, 56, 45-48.

- Frey, A., Hartig, J. & Moosbrugger, H. (2009). Effekte des adaptiven Testens auf die Motivation zur Testbearbeitung am Beispiel des Frankfurter Adaptiven Konzentrationsleistungs-Tests. *Diagnostica*, 55, Heft 1, 20-28.
- Funke, J. & Vaterrodt-Plünnecke, B. (2004). *Was ist Intelligenz?* München: Beck.
- Gage, N. L. & Berliner, D. C. (1986). *Pädagogische Psychologie* (4., völlig neu bearbeitete Auflage). München: PVU.
- Gerhard, A.-K. (2005). *Autonomie und Nähe - Individuationsentwicklung Jugendlicher im Spiegel familiärer Identifikation*. Weinheim, München: Juventa Verlag.
- Gneezy, U. & Rustichini, A. (2000). Pay enough or don't pay at all. *The Quarterly Journal of Economics*, Vol. 115, Issue 3, 791-810.
- Green, P. (2004). Testmotivation und ihre Messung. *Report Psychologie*, 29 (5), 303-308.
- Green, P., Rohling, M. L., Lees-Haley, P. R. & Allen, L. M. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury*, 15, 1045-1060.
- Guay, F., Vallerand, R. F. & Blanchard, C. (2000). On the Assessment of Situational Intrinsic and Extrinsic Motivation: The Situational Motivation Scale (SIMS). *Motivation and Emotion*, Vol. 24, No. 3, 175-213.
- Häcker, H. & Stapf, K.-H. (2009). *Psychologisches Wörterbuch* (15. überarbeitete und erweiterte Auflage). Bern: Hans Huber.
- Heckhausen, H. (1963). *Hoffnung und Furcht in der Leistungsmotivation*. Meisenheim am Glan: Anton Hain.
- Heckhausen, H. (1972). Die Interaktion der Sozialisationsvariablen in der Genese des Leistungsmotivs. In C. F. Graumann (Hrsg.), *Handbuch der Psychologie - Band 7/2* (S. 955-1019). Göttingen: Hogrefe.
- Heckhausen, H. (1975). Fear of failure as a self-reinforcing motive system. In I.G. Sarason & C. Spielberger (Eds.), *Stress and anxiety* (Vol. II), (S. 117-128). Washington: Hemisphere.
- Heckhausen, H. (1989). *Motivation und Handeln* (2. Auflage). Berlin: Springer.
- Heckhausen, J. & Heckhausen, H. (2006). *Motivation und Handeln* (3. überarbeitete und aktualisierte Auflage). Heidelberg: Springer Medizin Verlag.
- Heider, F. (1958). *The psychology of interpersonal relations*. New York: Wiley (deutsch: Psychologie der interpersonalen Beziehungen. Stuttgart: Klett, 1977).
- Helmke, A. (1992). Determinanten der Schulleistung: Forschungsstand und Forschungsdefizit. In G. Nold (Hrsg.), *Lernbedingungen und Lernstrategien. Welche Rolle spielen kognitive Verstehtensstrukturen?* (S. 23-34), Tübingen: Narr.

- Helmke, A. & Weinert, F. E. (1997). Bedingungsfaktoren schulischer Leistungen. In F. E. Weinert (Hrsg.), *Psychologie des Unterrichts und der Schule* (S. 71-176). Göttingen: Hogrefe.
- Horn, W. (1962, 1983). *Leistungsprüfsystem – LPS. Handanweisung für die Durchführung, Auswertung und Interpretation*. Göttingen: Hogrefe.
- Horner, M. S. (1968). *Sex differences in achievement motivation and performance in competitive and non-competitive situations*. University of Michigan: Unveröffentlichte Dissertation.
- Jacobs, B. (2007). *Geld und Noten als extrinsische Motivatoren zur Verbesserung kognitiver Leistungen*. (Online verfügbar: http://www.phil.uni-sb.de/~jakobs/wwwartikel/extrinsische_motivierung/index.html)
- Jacobs, B. (2009). *Geld zur Steigerung der Testmotivation- und Leistung in einer unbenoteten Abschlussklausur?* (Online verfügbar: <http://psydok.sulb.uni-saarland.de/volltexte/2009/2360/pdf/cash4quiz.pdf>)
- Jäger, A. O. (1973). *Dimensionen der Intelligenz* (3. Auflage). Göttingen: Hogrefe.
- Jensen, A. R. (1998). The g factor and the design of education. In R. J. Sternberg & W. M. Williams (Eds.), *Intelligence, instruction, and assessment: Theory into practice* (S. 111-131). Mahwah, NJ: Lawrence Erlbaum.
- Keil, A. (2009). *Die Beziehung von Selbständigkeit, Schulklassenklima und verbaler Intelligenz zur Leistungsmotivation von Hauptschülern*. Universität Regensburg: Unveröffentlichte Diplomarbeit.
- Keller, J. A. (1981). *Grundlagen der Motivation*. München: Urban & Schwarzenberg.
- Kieffer, D. A. & Goh, D. S. (1981). The effect of individually contracted incentives on intelligence test performance of middle- and low-SES children. *Journal of Clinical Psychology*, Vol. 37(1), 175-179.
- Köller, O. & Baumert, J. (2002). Entwicklung schulischer Leistungen. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie* (S. 756-786). Weinheim: Beltz.
- Kreuzpointner, L. (2010). *Bedingung für die Äquivalenz von Papier-Bleistift-Version und Computerversion bei Leistungstests*. Universität Regensburg: Dissertation.
- Krohne, H. W. & Hock, M. (2007). *Psychologische Diagnostik – Grundlagen und Anwendungsfelder*. Stuttgart: Kohlhammer.
- Kühn, R. (1987). *Bedingungen für Schulerfolg – Zusammenhänge zwischen Schülermerkmalen, häuslicher Umwelt und Schulnoten*. Göttingen: Hogrefe.
- Langens, T. A., Schmalt, H.-D. & Sokolowski, K. (2005). Motivmessung: Grundlagen und Anwendungen. In R. Vollmeyer & J. C. Brunstein (Hrsg.), *Motivationspsychologie und ihre Anwendung* (S. 70-89). Stuttgart: Kohlhammer.

- Lewin, K. (1938). *The conceptual representation and the measurement of psychological forces*. Durham, N. C.: Duke University Press.
- Lienert, G. A. & Raatz, U. (1998). *Testaufbau und Testanalyse* (6. Auflage). Weinheim: Beltz.
- Limbourg, M. (1998). Die Bedeutung von Familie und Freunde für Sicherheit und Gefahr im Kindes- und Jugendalter. *Vortrag bei dem XVIII Internationalen Symposium „Die Vermittlung von Sicherheit und Risiko in Familie und Schule“ in Köln, 11 1998, Gesellschaft für Sicherheitswissenschaft*. (Online verfügbar: <http://www.uni-due.de/~qpd402/alt/texte.ml/Familie.html>)
- Lohaus, A., Vierhaus, M. & Maass, A. (2010). *Entwicklungspsychologie des Kinder- und Jugendalters*. Heidelberg: Springer.
- Lowell, E. L. (1952). The effect of need for achievement on learning and speed of performance. *Journal of Psychology*, 33, 31-40.
- Lübke, J. (2007). *Die Auswirkung von Familienklima, Anerkennung in der Klasse und Leistung auf die Leistungsmotivation von Realschülern*. Universität Regensburg: Unveröffentlichte Diplomarbeit.
- Lukesch, H. (in Vorb.). *RLMI-K/J Regensburger Leistungs-Motiv-Inventar für Kinder und Jugendliche*. In Vorbereitung.
- Lukesch, H. (2006). *Einführung in die Pädagogische Psychologie*. Regensburg: Roderer.
- Lukesch, H. & Peters-Häderle, K. (2007). *RLMI-E Regensburger Leistungs-Motiv-Inventar für Erwachsene Manual*. Göttingen: Hogrefe.
- Mayes, S. D., Calhoun, S. L., Bixler, E. O. & Zimmerman, D. N. (2009). IQ and neuropsychological predictors of academic achievement. *Learning and Individual Differences*, Vol.19 (2), 238-241.
- McClelland, D. C., Atkinson, J. W., Clark, R. A. & Lowell, E. L. (1953). *The achievement motive*. New York: Appleton-Century-Crofts.
- McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities: Past, present and future. In D. P. Flanagan & P. L. Harison (Eds.) *Contemporary intellectual assessment: Theorie, tests, and issues* (S. 136-202). New York: Guilford Press.
- Meer, E. van der (1998). Intelligenz als Informationsverarbeitung. In E. Roth (Hrsg.), *Intelligenz. Grundlagen und neuere Forschung* (S. 161-184). Berlin: Kohlhammer.
- Merten, T. & Brockhaus, R. (2004). Ein Paradigmenwechsel in der Neuropsychologie. *Report Psychologie*, 29, 5, 302.
- Murray, H. A. (1938). *Explorations in personality*. New York: Oxford University Press.

- Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J. & Urbina, S. (1996). Intelligence: Knowns and Unknowns. *American Psychologist*, 51 (2), 77-101.
- Neugebauer, L. (1997). *Unternehmertum in der Unternehmung: Ein Beitrag zur Intrapreneurship-Diskussion*. Göttingen: Vandenhoeck & Ruprecht.
- Oerter, R. (1975). *Moderne Entwicklungspsychologie* (15. Auflage). Donauwörth: Auer Verlag.
- Petermann, F. & Winkel, S. (2007). *Fragebogen zur Leistungsmotivation für Schüler der 7. bis 13. Klasse - FLM 7-13*. Frankfurt/Main: Harcourt.
- Peters-Häderle, K. (2006). *Erfolgsschmerz und Leistungsangst bei Kindern. Eine Trainingsstudie*. Universität Regensburg: Dissertation.
- Pohlmann, B., Möller, J. & Streblow, L. (2005). Bedingungen leistungsbezogenen Verhaltens im Sportunterricht. *Zeitschrift für Sportpsychologie*, 12 (4), 127-134.
- Rauner, F., Haasler, B., Heinemann, L. & Grollmann, P. (2009). *Messen beruflicher Kompetenzen: Band II – Ergebnisse Komet 2008, Band 2*. Berlin: LIT Verlag.
- Raven, J. C. (2002). *Raven-Matrizen-Test / Coloured progressive matrices*. Frankfurt: Swets Test Services.
- Reeve, C. L. & Lam, H. (2007). Consideration of g as a common antecedent for cognitive ability test performance, test motivation, and perceived fairness. *Intelligence* Volume 35, Issue 4, 347-358.
- Rheinberg, F. (2006). *Motivation* (6. überarbeitete und erweiterte Auflage). Stuttgart: Kohlhammer.
- Rheinberg, F. (2004). *Motivationsdiagnostik. Kompendien. Psychologische Diagnostik. Band 5*. Göttingen: Hogrefe.
- Rheinberg, F., Vollmeyer, R. & Burns, B. D. (2001). *FAM: Ein Fragebogen zur Erfassung aktueller Motivation in Lern- und Leistungssituationen – (Langversion: 2001)*. (Online verfügbar: <http://psych-server.psych.uni-potsdam.de/people/rheinberg/messverfahren/FAMLangfassung.pdf>)
- Rost, D. H. (2009). *Intelligenz. Fakten und Mythen*. Weinheim, Basel: Beltz Verlag.
- Roth, E. (1998). Geschichte der Intelligenzforschung. In E. Roth (Hrsg.), *Intelligenz. Grundlagen und neuere Forschung* (S. 21-36). Berlin: Kohlhammer.
- Rudolph, U. (2009). *Motivationspsychologie* (2., vollständig überarbeitete Auflage). Weinheim, Basel: Beltz Verlag.
- Salgado, J. F., Remeseiro, C. & Iglesias, M. (1996). Personality and Test Taking Motivation. *Psicothema*, 8,(3), 553-562.

- Sanchez, R. J., Truxillo, D. M. & Bauer, T. N. (2000). Development and examination of an expectancy-based measure of test-taking motivation. *Journal of Applied Psychology*, 85, 739–750.
- Sarason, I. G. (1961). The effects of anxiety and threat on the solution of a difficult task. *Journal of Abnormal and Social Psychology*, 62 (1), 165-168.
- Sarason, I. G. (1963). Test anxiety and intellectual performance. *Journal of Abnormal and Social Psychology*, 66 (1), 73-75.
- Schmalt, H.-D. (1976). *Methoden der Leistungsmotivmessung*. Göttingen: Hogrefe.
- Schmalt, H.-D. (1976). *Das LM-GITTER. Ein objektives Verfahren zur Messung des Leistungsmotivs bei Kindern. Handanweisung*. Göttingen: Hogrefe.
- Schmalt, H.-D. (1986). *Motivationspsychologie*. Stuttgart: Kohlhammer.
- Schmalt, H.-D. (1999). Assessing the achievement motive using the grid technique. *Journal of Research in Personality*, 33, 109-130.
- Schmalt, H.-D. (2005). Validity of a short form of the Achievement-Motive Grid (AMG-S): Evidence for the three-factor structure, emphasizing active and passive forms of fear of failure. *Journal of Personality Assessment*, 84, 172-184.
- Schmalt, H.-D. & Meyer, W.-U. (1976). *Leistungsmotivation und Verhalten*. Stuttgart: Ernst Klett Verlag.
- Schmidbauer, E. (2008). *Analyse von Leistungsmotivation bei Realschülern*. Universität Regensburg: Unveröffentlichte Diplomarbeit.
- Schneider, K., Gallitz, H. & Meise, C. (1973). *Motivationsforschung. Motivation unter Erfolgsrisiko. Band 1*. Göttingen: Hogrefe.
- Segal, C. (2006). *Motivation, Test Scores, and Economic Success*. (Online verfügbar: <http://www.econ.upf.edu/docs/seminars/segal.pdf>)
- Sowa, D. (2009). *Leistungsmotivation und Selbstkonzept – Eine Validierungsstudie zum RLMI-K/J*. Universität Regensburg: Unveröffentlichte Diplomarbeit.
- Spearman, C. (1904). „General Intelligence“, objectively determined and measured. *American Journal of Psychology*, 15, 201-293.
- Stadler, M. (2010). *Die Beeinflussbarkeit der Intelligenztestleistung durch Testdurchführungssituation, Leistungsmotiv und implizite Intelligenztheorie*. Universität Regensburg: Unveröffentlichte Diplomarbeit.
- Stapf A. (2003). *Hochbegabte Kinder: Persönlichkeit, Entwicklung, Förderung*. München: C.H. Beck Verlag.
- Stern, W. (1912). *Die psychologischen Methoden der Intelligenzprüfung und deren Anwendung an Schulkindern*. Leipzig: Barth.

- Sundre, D.L. (2007). *The Student Opinion Scale (SOS). A measure of examinee motivation. Test Manual.* (Online verfügbar: http://www.jmu.edu/assessment/resources/resource_files/sos_manual.pdf)
- Sundre, D. L., & Kitsantas, A. (2004). An exploration of the psychology of the examinee: Can examinee self-regulation and test-taking motivation predict consequential and non-consequential test performance? *Contemporary Educational Psychology*, 29 (1), 6-26.
- Di, W. B. (1983). Self-verification: Bringing social reality into harmony with the self. In J. Suls & A. G. Greewald (Eds.), *Psychological perspectives on the self* (Vol. 2), (S. 33-66). Hillsdale, N.J.: Erlbaum.
- Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.
- Vollmeyer, R. & Brunstein, J. (2005). *Motivationspsychologie und ihre Anwendung*. Stuttgart: Kohlhammer.
- Wechsler, D. (1964). *Die Messung der Intelligenz Erwachsener*. Stuttgart: Huber.
- Weiner, B. (1994). *Motivationspsychologie* (3. Auflage). Weinheim: Beltz.
- Weiner, B., Frieze, I., Kukla, A., Reed, L., Rest, S. & Rosenbaum, R. M. (1971). Perceiving the causes of success and failure. In E.E. Jones, D.E. Kanouse, H.H. Kelley, R.E. Nisbett, S. Valins & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (S. 95-120). Morristown: General Learning Press.
- Wigfield, A. & Eccles, J. S. (2000). Expectancy-Value Theory of Achievement Motivation. *Contemporary Educational Psychology*, 25, 68-81.
- Wise, S. L. & DeMars, C. E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment*, 10, 1-17.
- Wolf, L. F. & Smith, J. K. (1995). The consequence of consequence: Motivation, anxiety, and test performance. *Applied Measurement in Education*, 8 (3), 227 – 242.
- Wolf, L. F., Smith, J. K. & Birnbaum, M. E. (1995). Consequence of performance, test motivation, and mentally taxing items. *Applied Measurement in Education*, 8(4), 341-351.

8. Erklärung

Ich habe die Arbeit selbstständig verfasst, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt und bisher keiner anderen Prüfungsbehörde vorgelegt.

Regensburg, den 24.01.2012

Ines Giermann

9. Anhang

9.1 Anhang A – Erhebungsmaterial

A-1 Elternanschreiben und Einverständniserklärung - Beispiel Gymnasium

Sehr geehrte Eltern und Erziehungsberechtigte,

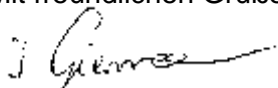
im Rahmen meiner Diplomarbeit an der Universität Regensburg möchte ich den Zusammenhang von Leistungsmotiven, Testmotivation und den Ergebnissen in einem Leistungstest bei Schülern untersuchen. Dabei sollen zur Erhebung des Leistungsmotivs das Leistungsmotivgitter von Prof. Schmalt und das Regensburger Leistungsmotivinventar von Prof. Lukesch eingesetzt werden. Die Testmotivation möchte ich mittels eines selbst konzipierten Fragebogens und die kognitive Leistungsfähigkeit mithilfe der revidierten Form des Leistungsprüfsystems ermitteln. Um die Schüler etwas genauer beschreiben zu können, möchte ich zudem die Schulnoten Ihres Kindes aus dem letzten Schuljahr sowie einige soziodemografischen Merkmale erfassen.

Diese Untersuchung hat weder einen Einfluss auf die Schulleistung noch findet dadurch eine, für die Schule relevante, Leistungsbeurteilung Ihres Kindes statt. Sie dient lediglich wissenschaftlichen Erkenntnissen über den Zusammenhang von Leistungsmotivation und Leistungsfähigkeit.

Die gesamte Untersuchung wird während der Unterrichtszeit stattfinden und insgesamt etwa eine Doppelstunde in Anspruch nehmen. Der genaue Zeitpunkt wird mit der Schulleitung und den Lehrkräften besprochen. Die Daten, die ich dabei erhebe, werden selbstverständlich *vertraulich behandelt* und nicht nach außen gelangen. Es wird lediglich ein individueller Code benötigt, um die unterschiedlichen Fragebögen einander zuordnen zu können.

Damit Ihr Kind an unserer Untersuchung teilnehmen darf, möchte ich Sie bitten, die unten angehängte *Einverständniserklärung* zu unterschreiben. Vielen Dank im Voraus für Ihre Mitarbeit!

Mit freundlichen Grüßen



Ines Giermann

Die Einverständniserklärung bitte bei der Lehrkraft abgeben.

Ich bin damit einverstanden / nicht einverstanden,

dass mein(e) Sohn / Tochter _____ an einer Untersuchung, über den Zusammenhang von Leistungsmotiven, Testmotivation und den Ergebnissen in einem Leistungstest bei Schülern, teilnimmt. Die Angaben meines Kindes und deren Auswertung unterliegen dem Datenschutz.

Ort, Datum

Unterschrift eines Erziehungsberechtigten

A-2 Soziodemographische Daten und Schulnoten - Beispiel Gymnasium

VP-Code: __ - __ - __

Soziodemographische Daten und Schulnoten

Bitte trage zunächst, bevor Du mit dem Ausfüllen beginnst, Deinen Versuchspersonencode, rechts oben auf diesem Blatt ein. Dieser Versuchspersonencode besteht aus den ersten beiden Buchstaben des Vornamens Deines Vaters, den ersten beiden Buchstaben des Vornamens Deiner Mutter und der Monatszahl Deines Geburtstages.

Soziodemographische Daten:

Geburtsjahr der Schülerin bzw. des Schülers: _____

Geschlecht:

☐ weiblich ☐ männlich

Besuchte Schulart und Jahrgangsstufe im Schuljahr 2009/2010:

☐ __. Klasse ☐ Hauptschule ☐ Realschule ☐ Gymna-

sium

Schulnoten aus dem Schuljahr 2009/2010:

Hauptfächer:	Note
Deutsch	
Englisch	
2. Fremdsprache	
Mathe	
Physik	
Chemie	

Nebenfächer:	Note
Erdkunde	
Geschichte	
Musik	
Biologie	
Religionslehre / Ethik	
Wirtschaft und Recht	
Sport	

A-3 Ergebnisrückmeldung an die Klassen

Ergebnisse des Leistungstests im Rahmen meiner Diplomarbeit

Liebe Schüler und Schülerinnen, liebe Eltern und Schulleiter,
die Ergebnisse meiner Untersuchung, welche ich im Januar und Februar 2011 durchgeführt habe, können nun eingesehen werden.

Vorab wollte ich mich nochmals bei allen Schülern und Schülerinnen, die an der Untersuchung teilgenommen haben, für die Teilnahme bedanken, aber auch bei den Eltern, die die Teilnahme ihres Kindes an dieser Untersuchung ermöglicht haben. Zudem gilt mein Dank auch den Schulleitern, die meiner Untersuchung offen gegenüberstanden und den betroffenen Lehrern, danke ich für ihre Mithilfe und Unterstützung.

Die Ergebnisse können unter folgendem Link eingesehen werden:

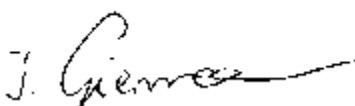
<http://www-cgi.uni-regensburg.de/Fakultaeten/Psychologie/Lukesch/lps.php>

Link kurz: www.ur.de/lukesch/lps

Um die Ergebnisse abrufen zu können, brauchst Du lediglich Deinen Versuchspersonencode. Nochmal zur Erinnerung: Dieser Versuchspersonencode besteht aus den ersten beiden Buchstaben des Vornamens Deines Vaters, den ersten beiden Buchstaben des Vornamens Deiner Mutter und der Monatszahl Deines Geburtstages. Da einige Versuchspersonencodes identisch waren, wurden, um diese auseinanderhalten zu können, noch zwei zusätzliche Kürzel mit aufgenommen: Schule: G für Gymnasium, R für Realschule und M für Mittelschule; Jahrgangsstufe: 9 für 9.Klasse und 10 für 10.Klasse. Somit ergibt sich beispielsweise folgender Code: Der Schüler hat am 28.**03**. Geburtstag, die Eltern heißen **F**lorian und **A**ndrea und er geht aufs **G**ymnasium in die **9**.Klasse: Der Code wäre hier: FLAN03G9.

Vielen Dank nochmals für die Teilnahme an meiner Untersuchung, auch im Namen meines Diplomarbeitbetreuer Ludwig Kreuzpointner.

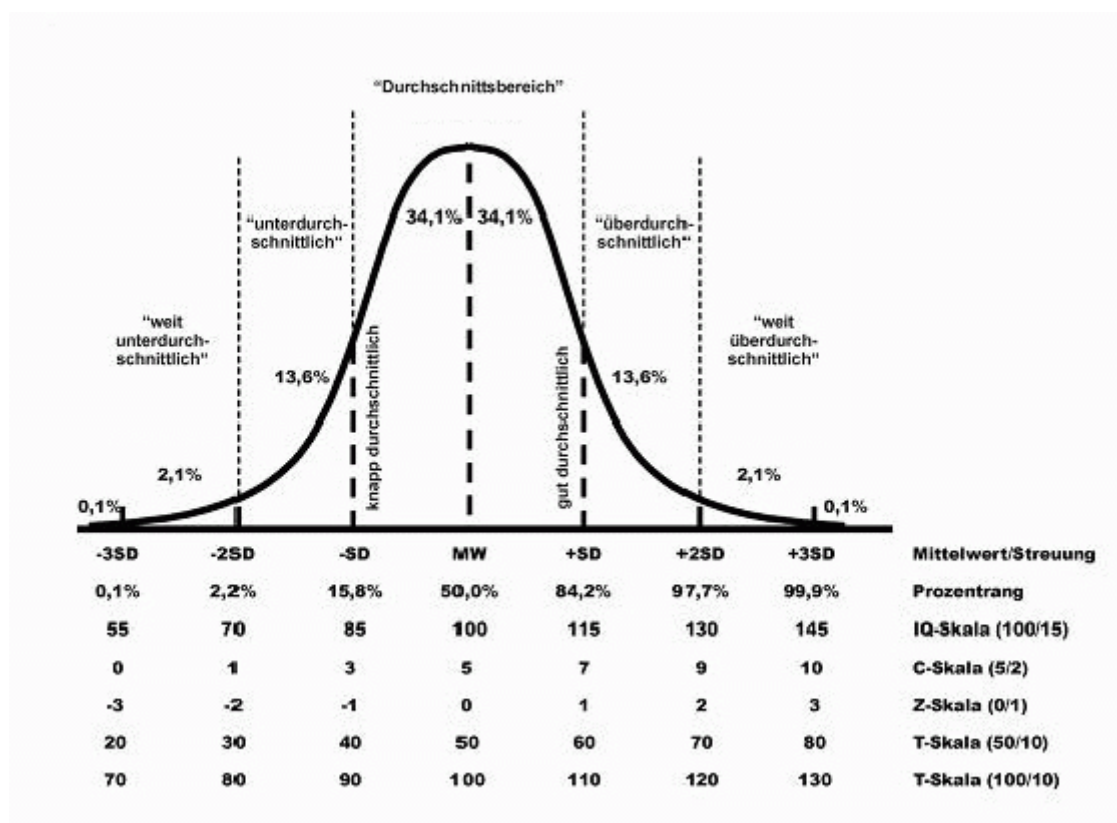
Mit freundlichen Grüßen


Ines Giermann

A-4 Ergebnissrückmeldung an die Schüler/-innen mithilfe eines Internetlinks

Die Ergebnisse Deines LPS-Tests kannst Du in der Tabelle unten nachlesen. Doch zuvor noch eine kurze Erklärung zum IQ-Wert und dazu, was die Daten in der Tabelle bedeuten.

Ein (IQ)Wert von 100 bedeutet, dass Du eine durchschnittliche Leistung erbracht hast. Die Grafik veranschaulicht die Verteilung und Bedeutung des IQ-Wertes.



Falls Du in einem der Tests einen deutlich abweichenden Wert haben solltest (z.B. 70), kann dies mehrere Ursachen haben:

- Die getestete Eigenschaft ist bei Dir nicht so gut ausgeprägt, wie bei den anderen Teilnehmern (Beachte, dass jedes Ergebnis in Bezug zu den Ergebnissen entweder Deiner Mitschüler, Deiner Schule oder aller teilnehmenden Schulen gesetzt wurde!).
- Du hast Dir keine besondere Mühe bei der Testung gegeben.
- Ich habe einen Fehler bei der Auswertung gemacht.
- Der Test ist schlecht.
- ...

Zur Erinnerung:

Der Test den Du bearbeitet hast bestand aus elf Teilaufgaben, wobei jede einzelne einen unterschiedlichen Bereich der Intelligenz erfasst. Hier eine Übersicht:

Test Inhalt

- 1 Allgemeinwissen (falschen Buchstaben finden)
- 2 Sprachliches Denken (ersten Buchstaben finden)
- 3 Logisches Denken (Figurenreihen)
- 4 Logisches Denken (Zahlenreihen)
- 5 Logisches Denken (Buchstabenreihen)
- 6 Mentale Rotation (spiegelverkehrtes Zeichen finden)
- 7 Räumliches Vorstellen (Zahl der Flächen)
- 8 Strukturen erkennen (Muster einpassen)
- 9 Sorgfalt (0, 1, 2 zählen)
- 10 Aufmerksamkeit (Fehler finden)
- 11 Konzentration (Zahlen addieren)

Skalen aus den einzelnen Tests zusammengefasst:

- Gc Kristalline Intelligenz (Kulturelle Fähigkeit)
 Gf Fluide Intelligenz (Allgemeine Fähigkeit, schlussfolgernd zu denken)
 Gv Visuell-räumlich Fähigkeiten
 Gs Kognitive Schnelligkeit
 G Allgemeine Intelligenz

Die Abkürzungen findest Du auch in Deiner Ergebnistabelle wieder. Zudem siehst Du in der Tabelle Deine Leistungen im Vergleich:

1. zu Deinen Mitschülern in den jeweiligen Bereichen (Klasse).
2. zu Deiner Schule in den jeweiligen Bereichen (Schule).
3. zur Gesamtstichprobe (=alle Schulen: Gymnasium, Realschule und Mittelschule) in den jeweiligen Bereichen.

Eine kurze Erklärung der eingetragenen Daten: IQ-LPS_Gc_Klasse = 100: gibt Deinen IQ-Wert im Bereich der kristallinen Intelligenz im Vergleich zur gesamten Klasse an. Ein Wert von 100 würde bedeuten, dass Du im Bereich der kristallinen Intelligenz eine durchschnittliche Leistung im Vergleich zu Deinen Mitschülern erbracht hast (siehe Grafik oben).

Code	IQ-LPS_Gc	IQ-LPS_Gf	IQ-LPS_Gv	IQ-LPS_Gs	IQ-LPS_G
	81	102	99	82	86

Anmerkungen: Auf die Darstellung der Ergebnisse im Vergleich zur Klasse und zur Schule wurde hier verzichtet. Diese wurden den Schüler/-innen auf der Internetseite ebenfalls mitgeteilt.

A-5 Fragebogen zur Erfassung der Testmotivation

Im Anschluss an diesen Fragebogen wirst Du einen Leistungstest bearbeiten, der aus elf verschiedenen Aufgabenteilen besteht. Durch diesen Test soll festgestellt werden, welche Aufgaben Dir leicht fallen und welche Dir eher weniger liegen.

Die Aussagen und Fragen in diesem Fragebogen beziehen sich fast ausschließlich auf den nachfolgenden Leistungstest. Bitte denke kurz über die Aussagen nach und kreuze an, wie sehr diese auf Dich zutreffen.

Es handelt sich um einen Fragebogen zu persönlichen Einstellungen. Daher gibt es keine richtigen oder falschen Antworten. Denke darüber nicht lange nach, sondern versuche, Deinem spontanen Eindruck zu folgen.

Bitte trage zunächst, bevor Du mit dem Fragebogen beginnst Deinen Versuchspersonencode rechts oben auf dem nächsten Blatt ein. Dieser Versuchspersonencode besteht aus den ersten beiden Buchstaben des Vornamens Deines Vaters, den ersten beiden Buchstaben des Vornamens Deiner Mutter und der Monatszahl Deines Geburtstages.

Fragebogen - vor dem Test

Aussagen zu Deinem allgemeinen Leistungsverhalten	stimme nicht zu			stimme voll zu	
	1	2	3	4	5
1. Ich bemühe mich darum, dass meine Leistungen besser sind als der Durchschnitt. (vA1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. Ich vergleiche meine Leistungen mit denen der anderen. (vA2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. Ich lege Wert darauf, dass meine Leistungen anerkannt werden. (vA3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. Ich bearbeite am liebsten Aufgaben, die mich herausfordern. (vA4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. Auch bei langweiligen Aufgaben bleibe ich bei der Sache. (vA5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. Wenn mir eine Aufgabe nicht gleich gelingt, versuche ich alles, um doch eine Lösung zu finden. (vA6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7. Beim Lernen oder Arbeiten lasse ich mich nicht unterbrechen. (vA7)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Aussagen und Fragen zum nachfolgenden Test

Aussagen zu Deinen grundsätzlichen Einstellungen	stimme nicht zu			stimme voll zu	
	1	2	3	4	5
1. In Tests mag ich es, mich neuen geistigen Herausforderungen zu stellen. (vB1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. Mit Tests bzw. Testsituationen habe ich bisher gute Erfahrungen gesammelt. (vB2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. Ich glaube, dass meine Klassenkameraden sehr motiviert sind, diesen Test zu bearbeiten. (vB3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. Meinen Eltern ist es wichtig, dass ich in diesem Test mein Bestes gebe. (vB4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Aufgabenwert: subjektiver Wert des Tests	stimme nicht zu			stimme voll zu	
	1	2	3	4	5
1. Es ist für mich persönlich wichtig, in diesem Test gut zu sein. (vC1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. Ich möchte in diesem Test zu den Besten gehören. (vC2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. Wahrscheinlich gibt es gute Gründe sich in diesem Test anzustrengen, aber ich persönlich sehe keinen. (vC3u)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. Bei Aufgaben wie dieser brauche ich keine Belohnung, sie machen mir auch so viel Spaß. (vC4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. Einen solchen Test würde ich auch in meiner Freizeit bearbeiten. (vC5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. Ich freue mich auf die Bearbeitung des Tests. (vC6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Bitte umblättern.

Aufgabenwert: subjektiver Wert des Tests	stimme nicht zu		stimme voll zu		
	1	2	3	4	5
7. Ich kann durch die Bearbeitung des Tests bestimmt etwas Nützliches lernen. (vC7)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8. Ich sehe nicht, was dieser Test mir bringen soll. (vC8u)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
9. Ich habe Angst, dass ich in diesem Test nicht gut sein werde. (vC9)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
10. Ich setze mich selbst unter Druck, damit ich in diesem Test nicht schlecht abschneide. (vC10)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
11. Anstelle dieses Tests würde ich lieber etwas anderes machen. (vC11u)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Erfolgserwartung	stimme nicht zu		stimme voll zu		
	1	2	3	4	5
1. Ich werde in diesem Test bestimmt gut abschneiden. (vD1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. So ein Test fällt mir bestimmt leicht. (vD2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. So ein Test liegt mir bestimmt nicht. (vD3u)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. Ich hoffe, dass mir die Aufgaben in diesem Test genauso leicht fallen wie meinen Klassenkameraden. (vD4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. Dieser Test wird sehr schwierig sein. (vD5u)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	keine Anstrengung		volle Anstrengung		
	0%	25%	50%	75%	100%
1. Wie viel Anstrengung wirst Du aufbringen, um die Aufgaben in diesem Test zu beantworten? (v1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	nicht motiviert			sehr motiviert	
	1	2	3	4	5
2. Wie motiviert bist Du, in diesem Test Dein Bestes zu geben? (v2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	nicht wichtig			sehr wichtig	
	1	2	3	4	5
3. Wie wichtig ist es für Dich, in diesem Test Dein Bestes zu geben? (v3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
	nicht gut			sehr gut	
	1	2	3	4	5
4. Wie gut wirst Du in diesem Test abschneiden? (v4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Fragebogen - nach dem Test

Aussagen und Fragen zum vorangegangenen Test

Aussagen zu Deinem Leistungsverhalten im vorangegangenen Test	stimme nicht zu			stimme voll zu	
	1	2	3	4	5
1. Ich hätte mich mehr anstrengen können. (nA1u)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. Ich habe in diesem Test durchgehend eine hohe Anstrengung aufgebracht. (nA2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. Während des Tests war ich vollkommen aufmerksam. (nA3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. Als ich diesen Test bearbeitet habe, war ich konzentriert und habe versucht alle Aufgaben richtig zu beantworten. (nA4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. Wenn mir eine Aufgabe nicht gleich gelungen ist, habe ich alles versucht, um doch noch eine Lösung zu finden. (nA5)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. Ich fand es herausfordernd, mich auch den schwierigen Aufgaben zu stellen. (nA6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7. Die leichten Aufgaben haben mir besser gefallen, da ich hier gleich eine Lösung gefunden habe. (nA7)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Aufgabenwert: subjektiver Wert des vorangegangenen Tests	stimme nicht zu			stimme voll zu	
	1	2	3	4	5
1. In diesem Test gut zu sein, war mir persönlich sehr wichtig. (nB1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. Ich möchte in diesem Test zu den Besten gehören. (nB2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. Bei Aufgaben wie dieser brauche ich keine Belohnung, sie machen mir auch so viel Spaß. (nB3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. Einen solchen Test würde ich auch in meiner Freizeit bearbeiten. (nB4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
5. Ich habe durch die Bearbeitung des Tests nichts Nützliches gelernt. (nB5u)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
6. Ich finde diesen Test, verglichen mit anderen Aktivitäten, sehr nützlich. (nB6)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
7. Ich habe mich selbst unter Druck gesetzt, damit ich in diesem Test nicht schlecht abschneide. (nB7)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
8. Anstelle dieses Tests hätte ich lieber etwas anderes gemacht. (nB8u)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Bitte umblättern.

Erfolgserwartung	stimme nicht zu				stimme voll zu
	1	2	3	4	5
1. Solche Art von Tests liegt mir einfach nicht. (nC1u)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. Ich schätze meine Leistungen in diesem Test durchweg positiv ein. (nC2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. Wenn ich Aufgaben in diesem Test schwierig fand, erging es meinen Klassenkameraden bestimmt ebenso. (nC3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. Dieser Test ist mir leicht gefallen. (nC4)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	nicht motiviert			sehr motiviert	
	1	2	3	4	5
1. Wie motiviert warst Du, Dein Bestes in diesem Test zu geben? (n1)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	interessiert mich nicht			interessiert mich sehr	
	1	2	3	4	5
2. Interessiert es Dich, wie Du im Vergleich zu den Anderen abgeschnitten hast? (n2)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	keine Anstrengung		volle Anstrengung		
	0%	25%	50%	75%	100%
3. Bei Dingen, die Dir persönlich sehr wichtig sind, strengst Du Dich auch einmal besonders an und gibst Dein Bestes (z.B. Hobbys). Wie sehr hast Du Dich im Vergleich dazu in diesem Test angestrengt? (n3)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Vielen Dank.

Anmerkungen. Die Itembezeichnungen nach jeder Aussage stehen nicht im Original

9.2 Anhang B – Skalen und Itemkennwerte der verwendeten Fragebögen

Tabelle B-1

Skalenkennwerte des RLMI-K/J ($n = 348$)

Hoffnung auf Erfolg	Itemzahl	<i>M</i>	<i>SD</i>	Schiefe	Kurtosis
Gesamt	12	39.86	7.38	-.10	-.16
Ausbildung	6	19.18	3.92	-.19	.10
Freizeit	6	20.38	4.30	-.26	-.20
Furcht vor Misserfolg	Itemzahl	<i>M</i>	<i>SD</i>	Schiefe	Kurtosis
Gesamt	12	29.79	7.95	.13	-.41
Ausbildung	6	15.16	4.32	.01	-.33
Freizeit	6	14.63	4.57	.23	-.40
Furcht vor Erfolg	Itemzahl	<i>M</i>	<i>SD</i>	Schiefe	Kurtosis
Gesamt	12	23.34	8.07	.75	.39
Ausbildung	6	11.25	4.29	.82	.36
Freizeit	6	12.09	4.37	.68	.18
Hoffnung auf Misserfolg	Itemzahl	<i>M</i>	<i>SD</i>	Schiefe	Kurtosis
Gesamt	12	23.03	7.65	.84	.59
Ausbildung	6	10.97	4.37	1.00	.90
Freizeit	6	12.07	4.06	.59	-.04

Anmerkungen. *M* = Mittelwert; *SD* = Standardabweichung

Tabelle B-2

Skalenkennwerte des Leistungsmotivgitters ($n = 322$)

	Itemzahl	<i>M</i>	<i>SD</i>	Schiefe	Kurtosis
Hoffnung auf Erfolg	18	8.51	4.04	-.13	-.46
Furcht vor Misserfolg - aktiv	18	9.52	4.36	-.34	-.55
Furcht vor Misserfolg - passiv	18	4.43	3.02	.71	.87

Anmerkungen. *M* = Mittelwert; *SD* = Standardabweichung

Tabelle B-3

Skalenkennwerte des Fragebogens zur Erfassung der Testmotivation (n = 334)

AW	Itemzahl	M	SD	Schiefe	Kurtosis	mittlere Trennschärfe
Gesamt	22	63.48	15.18	.01	-.98	.54
Vor dem Test	13	38.74	9.47	.04	-.95	.56
Nach dem Test	9	27.74	7.21	.11	-.99	.55
EW	Itemzahl	M	SD	Schiefe	Kurtosis	mittlere Trennschärfe
Gesamt	11	36.35	5.51	-.20	-.34	.37
Vor dem Test	6	20.69	3.54	-.16	-.14	.46
Nach dem Test	5	15.66	3.20	-.11	-.62	.31
EL	Itemzahl	M	SD	Schiefe	Kurtosis	mittlere Trennschärfe
Gesamt	12	38.21	6.05	-.18	-.71	.32
Einst.	4	11.43	2.56	.01	-.73	.18
ALV	8	26.77	4.46	-.27	-.64	.34
	Itemzahl	M	SD	Schiefe	Kurtosis	mittlere Trennschärfe
LV	8	28.40	5.73	-.48	-.68	.51

Anmerkungen. *M* = Mittelwert; *SD* = Standardabweichung; AW = Aufgabenwert; EW = Erfolgserwartung; EL = Einstellungen und Leistungen; Einst. = Einstellungen; ALV = Allgemeines Leistungsverhalten; LV = Leistungsverhalten im Test

Tabelle B-4

Itemkennwerte des Fragebogens zur Erfassung der Testmotivation ($n = 334$)

	Item	<i>M</i>	<i>SD</i>	Trennschärfe (Bereich)	Trennschärfe (gesamt)
Aufgabenwert vorher	vC1	3.21	1.12	.69	.66
	vC2	3.11	1.21	.53	.49
	vC3	3.20	1.25	.53	.55
	vC4	3.07	1.10	.54	.54
	vC5	2.41	1.33	.50	.51
	vC6	3.14	1.09	.63	.60
	vC7	3.23	1.03	.60	.56
	vC8	3.21	1.21	.60	.60
	vC9	2.09	1.15	.36	.36
	vC10	1.97	1.11	.42	.40
	vC11	3.26	1.31	.47	.44
	v2	3.57	.99	.61	.60
	v3	3.28	1.14	.68	.69
	Item	<i>M</i>	<i>SD</i>	Trennschärfe (Bereich)	Trennschärfe (gesamt)
Aufgabenwert nachher	nB1	2.87	1.21	.69	.65
	nB2	2.82	1.17	.62	.60
	nB3	2.84	1.23	.52	.53
	nB4	2.16	1.24	.57	.55
	nB5	2.73	1.34	.32	.35
	nB6	2.55	1.17	.62	.57
	nB7	2.47	1.21	.53	.48
	nB8	2.98	1.32	.45	.43
	n1	3.34	1.12	.59	.59
	Item	<i>M</i>	<i>SD</i>	Trennschärfe (Bereich)	Trennschärfe (gesamt)
Erfolgserw. vorher	vD1	3.17	.84	.50	.52
	vD2	3.33	.91	.60	.52
	vD3	3.67	.96	.54	.47
	vD4	3.51	1.07	.13	.15
	vD5	3.67	1.03	.39	.29
	v4	3.34	.76	.54	.51
	Item	<i>M</i>	<i>SD</i>	Trennschärfe (Bereich)	Trennschärfe (gesamt)
Erfolgserw. nachher	nC1	3.08	1.06	.35	.31
	nC2	2.92	.98	.44	.46
	nC3	3.47	1.10	.18	.23
	nC4	2.64	.95	.28	.31
	n2	3.55	1.31	.28	.21

Anmerkungen. *M* = Mittelwert; *SD* = Standardabweichung; Trennschärfe (Bereich): Skalen getrennt nach vor und nach dem Test; Trennschärfe (gesamt): gesamte Skala

Fortsetzung: Tabelle B-4

Itemkennwerte des Fragebogens zur Erfassung der Testmotivation (n = 334)

	Item	M	SD	Trennschärfe (Bereich)	Trennschärfe (gesamt)
Allg. Leistungsverhalten	vA1	3.51	1.05	.44	.39
	vA2	3.73	1.10	.13	.12
	vA3	3.57	.98	.31	.31
	vA4	2.89	1.03	.44	.50
	vA5	2.97	1.12	.28	.34
	vA6	3.38	1.04	.43	.39
	vA7	2.89	1.17	.30	.35
	v1	3.81	.92	.33	.36
	Item	M	SD	Trennschärfe (Bereich)	Trennschärfe (gesamt)
Einstellung	vB1	2.99	1.03	.14	.40
	vB2	3.23	.95	.13	.19
	vB3	2.53	1.08	.29	.25
	vB4	2.68	1.35	.15	.25
	Item	M	SD	Trennschärfe	
Leistungsverhalten	nA1	3.20	1.29	.29	
	nA2	3.50	1.15	.64	
	nA3	3.57	1.13	.68	
	nA4	3.97	1.10	.69	
	nA5	3.34	1.22	.49	
	nA6	3.41	1.19	.46	
	nA7	4.08	1.10	.03	
	n3	3.33	1.06	.57	

Anmerkungen. M = Mittelwert; SD = Standardabweichung; Trennschärfe (Bereich): Skalen getrennt nach vor und nach dem Test; Trennschärfe (gesamt): gesamte Skala

9.3 Anhang C – Verteilung der LPS-neu Subtests und des Gesamtwertes

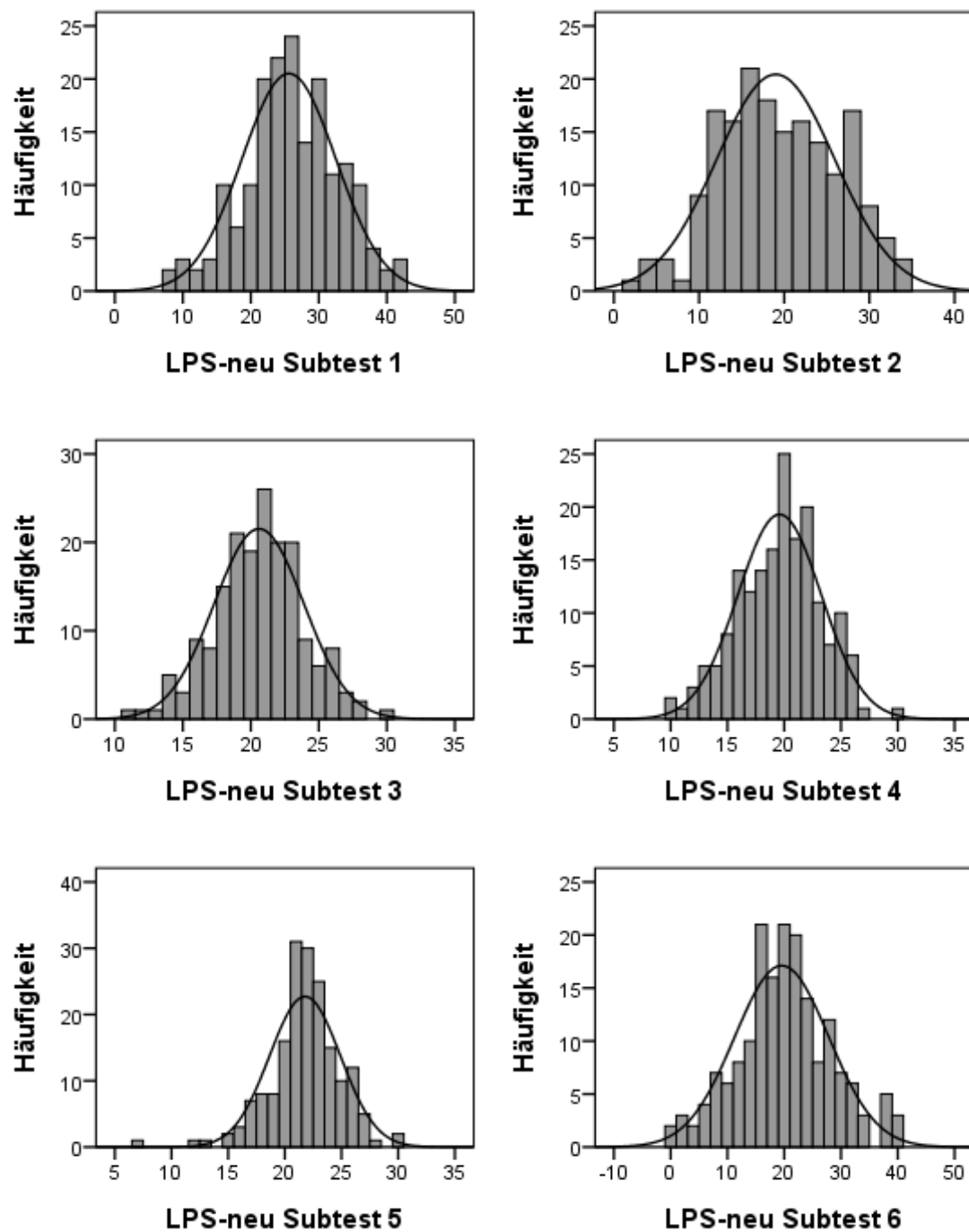
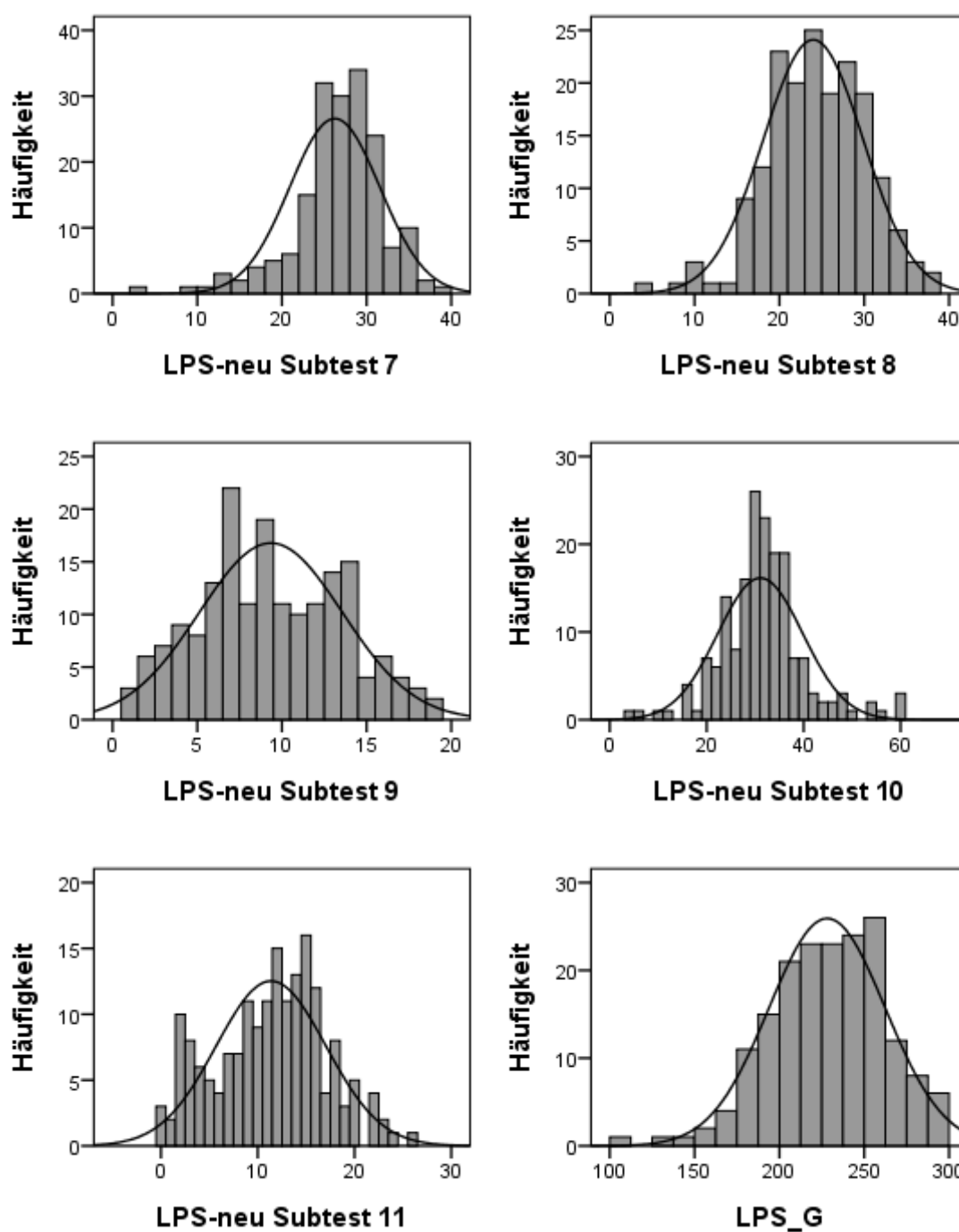


Abbildung C-1

Verteilung der LPS-neu Subtests und des LPS-Gesamtwertes für die Version A



Fortsetzung: Abbildung C-1

Verteilung der LPS-neu Subtests und des LPS-Gesamtwertes für die Version A

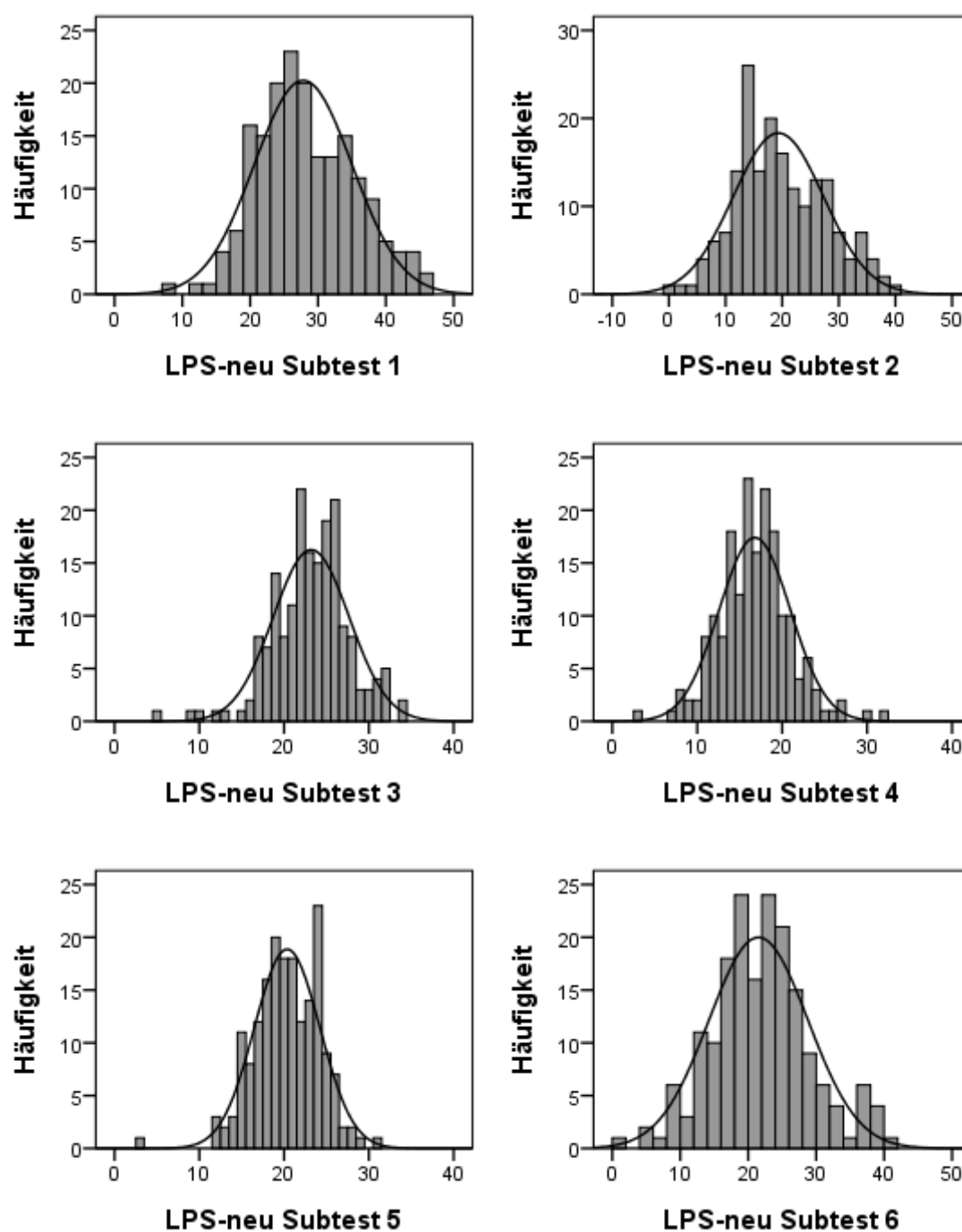
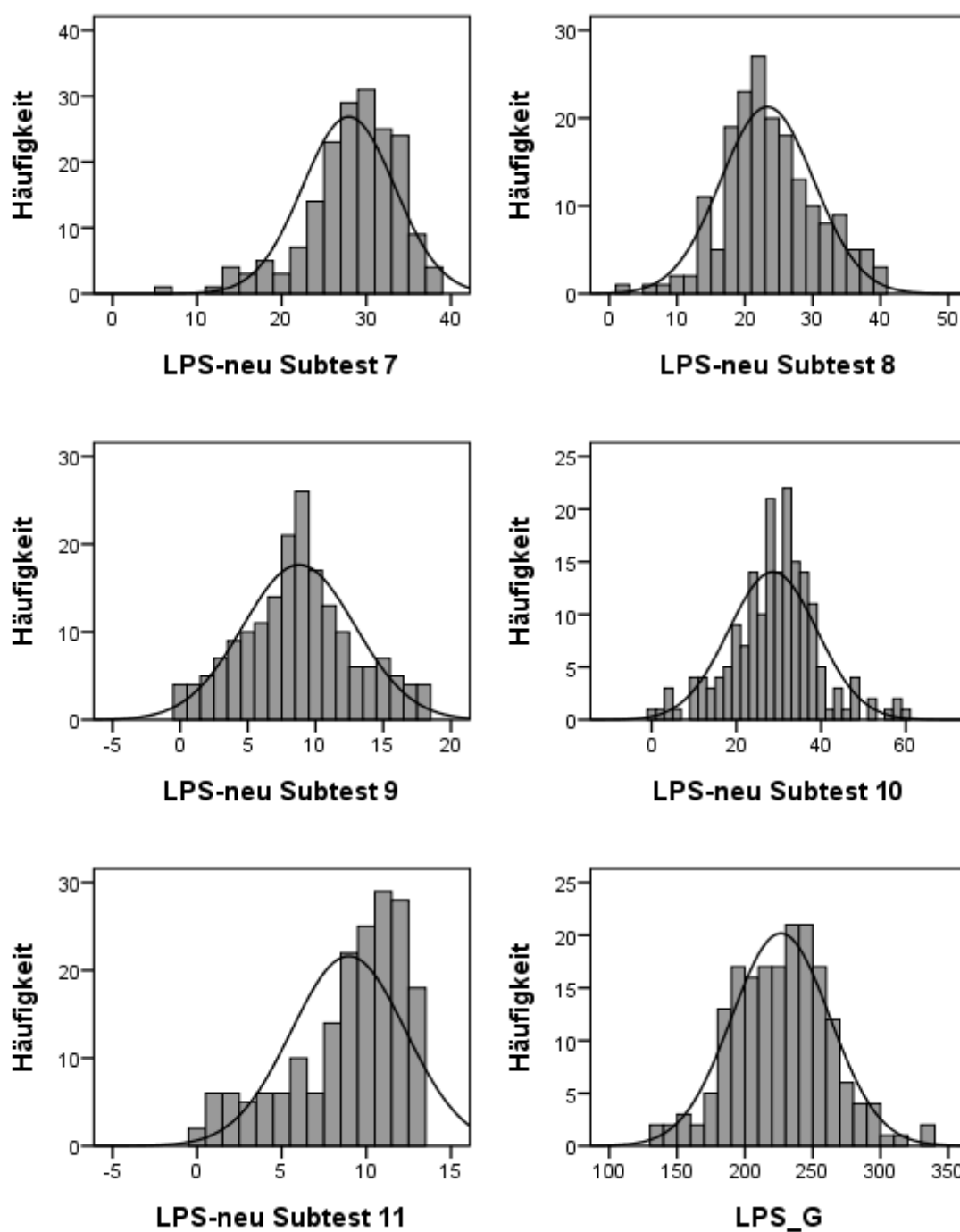


Abbildung C-2

Verteilung der LPS-neu Subtests und des LPS-Gesamtwertes für die Version B



Fortsetzung: Abbildung C-2

Verteilung der LPS-neu Subtests und des LPS-Gesamtwertes für die Version B

9.4 Anhang D – Strukturmodelle zwischen Testmotivation und LPS-neu

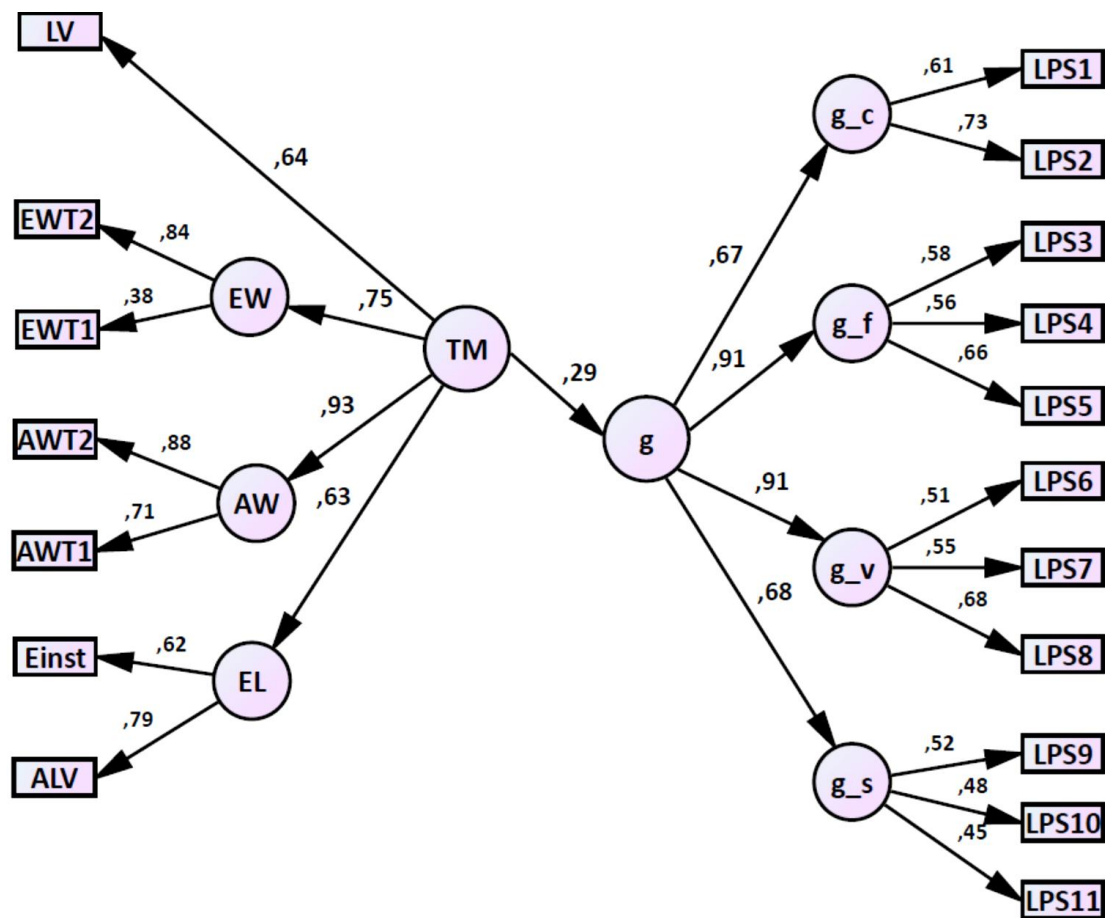


Abbildung D-1

Modell über den Zusammenhang zwischen der Testmotivation mit Skalen als beobachtbaren Variablen und dem LPS-neu (Version B)

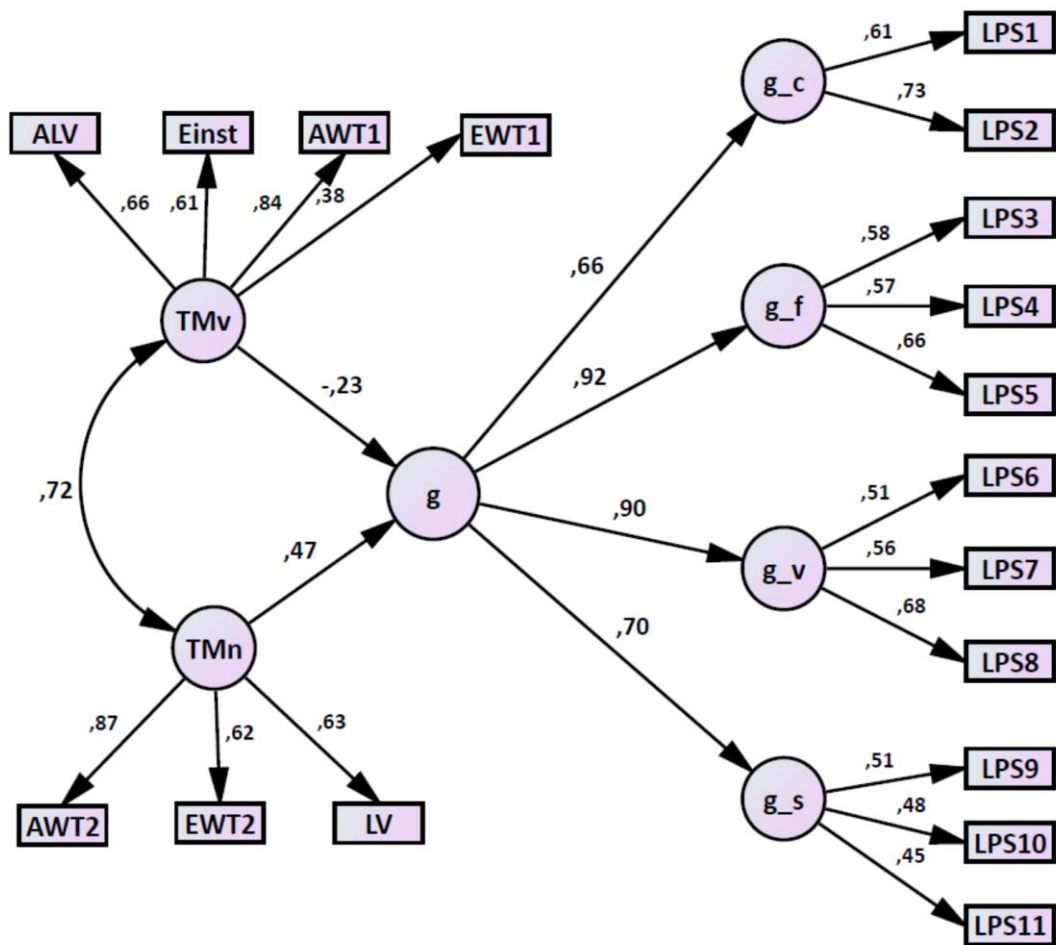


Abbildung D-2

Modell über den Zusammenhang zwischen der Testmotivation vor und nach dem Test und deren Einfluss auf das LPS-neu (Version B)

9.5 Anhang E – Zusammenhänge der Hauptfächernoten mit dem Gesamtwert und den Stratum II Komponenten des LPS-neu

Tabelle E-1

Zusammenhänge der Hauptfächernoten mit dem Gesamtwert und den Stratum II Komponenten des LPS-neu (Version A)

Mittelschule (n = 38)	Deutsch	Mathe	Englisch	
LPS_G	-.69**	-.04	-.39	
LPS_Gc	-.66**	.15	-.51*	
LPS_Gf	-.57**	-.28	-.26	
LPS_Gv	-.46*	-.24	-.19	
LPS_Gs	-.51*	.13	-.23	
Mädchenrealschule (n = 85)	Deutsch	Mathe	Englisch	
LPS_G	-.28	-.31*	.04	
LPS_Gc	-.39*	-.32*	-.22	
LPS_Gf	-.24	-.35*	-.15	
LPS_Gv	-.22	-.28	-.07	
LPS_Gs	-.00	-.03	.38*	
Knabenrealschule (n = 87)	Deutsch	Mathe	Englisch	4. Prüfungsfach
LPS_G	.03	-.12	.12	.21
LPS_Gc	-.06	-.01	-.03	.17
LPS_Gf	-.00	-.16	.17	.11
LPS_Gv	.05	-.12	.13	.23
LPS_Gs	.04	-.03	.05	.00
Gymnasium (n = 133)	Deutsch	Mathe	Englisch	2. Fremdsprache
LPS_G	.04	-.07	.01	.11
LPS_Gc	.06	.10	.09	.19
LPS_Gf	.17	-.01	.10	.17
LPS_Gv	-.07	-.26*	-.12	-.01
LPS_Gs	.02	-.00	.01	-.01

Anmerkungen. **) Die Korrelation ist auf dem .01 Niveau signifikant; *) Die Korrelation ist auf dem .05 Niveau signifikant

Tabelle E-2

Zusammenhänge der Hauptfächernoten mit dem Gesamtwert und den Stratum II Komponenten des LPS-neu (Version B)

Mittelschule (n = 38)	Deutsch	Mathe	Englisch	
LPS_G	-.44	-.69**	-.41	
LPS_Gc	-.39	-.50*	-.55*	
LPS_Gf	-.39	-.65**	-.32	
LPS_Gv	-.21	-.65**	-.19	
LPS_Gs	-.49*	-.46	-.35	
Mädchenrealschule (n = 85)	Deutsch	Mathe	Englisch	
LPS_G	-.25	-.06	-.08	
LPS_Gc	-.21	.24	-.11	
LPS_Gf	-.05	.02	.16	
LPS_Gv	-.20	-.11	-.12	
LPS_Gs	-.12	-.19	-.04	
Knabenrealschule (n = 87)	Deutsch	Mathe	Englisch	4. Prüfungsfach
LPS_G	-.08	-.14	-.20	-.21
LPS_Gc	-.13	.01	-.30	.05
LPS_Gf	.02	-.19	-.15	-.01
LPS_Gv	-.10	-.12	-.04	-.23
LPS_Gs	.01	-.14	-.08	-.33*
Gymnasium (n = 133)	Deutsch	Mathe	Englisch	2. Fremdsprache
LPS_G	.08	-.14	-.15	-.09
LPS_Gc	.11	-.00	-.02	-.16
LPS_Gf	.17	-.12	-.15	.08
LPS_Gv	-.05	-.12	-.06	-.05
LPS_Gs	.01	-.20	-.24*	-.12

Anmerkungen. **) Die Korrelation ist auf dem .01 Niveau signifikant; *) Die Korrelation ist auf dem .05 Niveau signifikant